RUTGERS
UNIVERSITY | NEW BRUNSWICK

Tutorial on Fairness of Machine Learning in Recommender Systems

Yunqi Li
Rutgers University
yunqi.li@rutgers.edu

Yingqiang Ge
Rutgers University
yingqiang.ge@rutgers.edu

Yongfeng Zhang
Rutgers University
yongfeng.zhang@rutgers.edu

Learning Objective

- Raise awareness on the importance of considering fairness in recommendation
- Get the background knowledge of fairness works in general machine learning
- Learn about the taxonomies of fairness concepts in recommendation
- Know about some datasets, evaluation protocols to assess fairness in recommendation
- Understand the challenges and opportunities of fairness research in recommendation

Outline

- Introduction:
  - Social Impact of Recommender System and Fairness
  - Motivation of Fairness in Recommender Systems
  - Relationship with AI Ethics
  - Beyond Ethics: a Utilitarian Perspective
- Fairness in Machine Learning:
  - Fairness in Classification
  - Fairness in Ranking
- Fairness in Recommendation:
  - Introduction
  - Taxonomy
  - Dataset and Evaluation
  - Challenge and Opportunity

RUTGERS

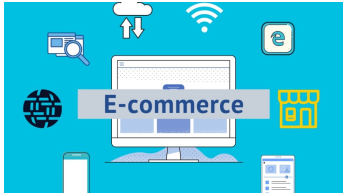
Outline

- Introduction:
  - Social Impact of Recommender System and Fairness
  - Motivation of Fairness in Recommender Systems
  - Relationship with AI Ethics
  - Beyond Ethics: a Utilitarian Perspective
- Fairness in Machine Learning:
  - Fairness in Classification
  - Fairness in Ranking
- Fairness in Recommendation:
  - Introduction
  - Taxonomy
  - Dataset and Evaluation
  - Challenge and Opportunity

Recommender Systems are Everywhere

- For example

E-commerce Systems



Product Recommendation

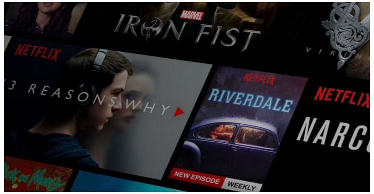
Social Networks

Newsfeed Recommendation
Friend Recommendation

Healthcare Systems

Doctor Recommendation
Patient-doctor matching

Online Entertainment Systems



Movie/Video Recommendation
Music Recommendation

Trip Planning Systems

Hotel Recommendation
Air Ticket Recommendation

Financial Applications

Investmemnt Recommendation

Cyber-Physical Systems

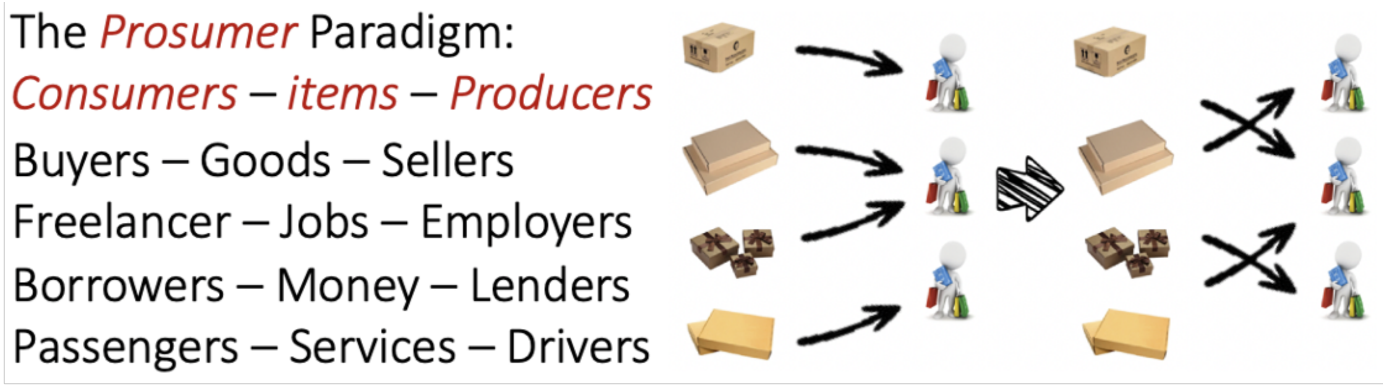
Driver Recommendation
Route Recommendation

Talent Recruiting Systems

Job Recommendation
Candidate Recommendation

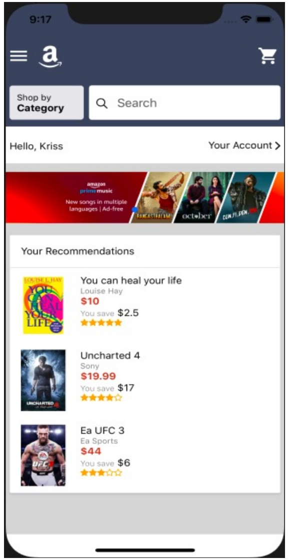
Social Impacts of Recommender Systems

- Recommender Systems are far more than just information seeking tools
  - They control how resources are allocated among differnet parties
    - Resources can be exposure opportunies, products, jobs, information, etc.
    - Usually RS works in two-sided markets/environments [1]

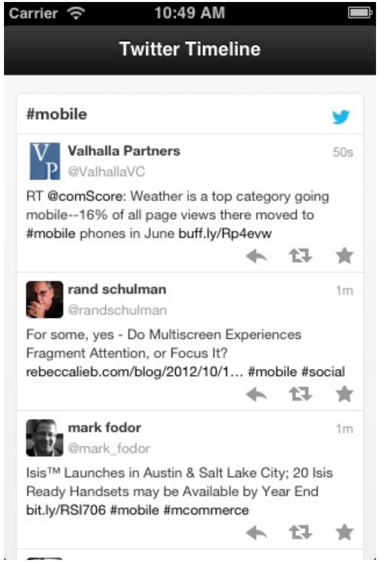


[1] Y. Zhang, Q. Zhao, Y. Zhang, D. Friedman, M. Zhang, Y. Liu, S. Ma. Economic Recommendation with Surplus Maximization. WWW 2016.

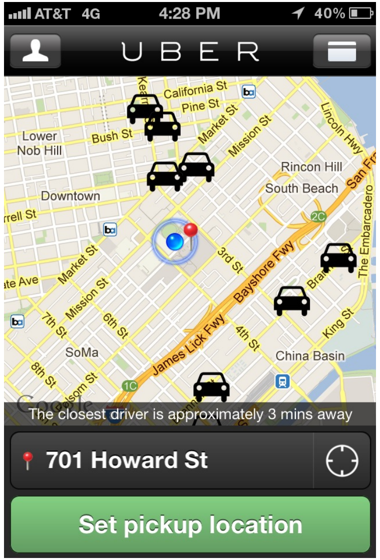
Why Fairness in RecSys? Resources Could be Limited



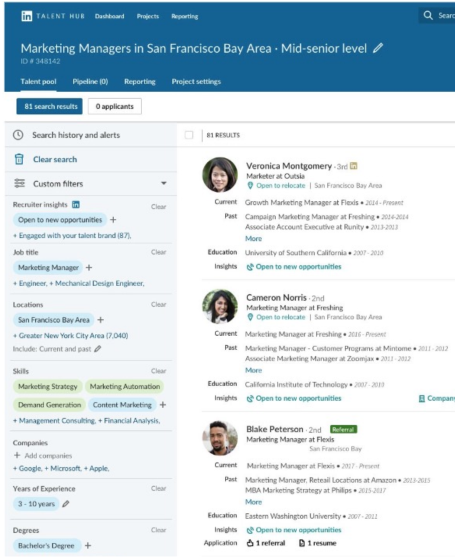
Recommendation slot positions are limited, which producers' items should be recommended and get the exposure opportunity to users?



User attention is a limited resource, whose twite should get exposure on the timeline?



Passengers are limited, which driver should get the task and make money?

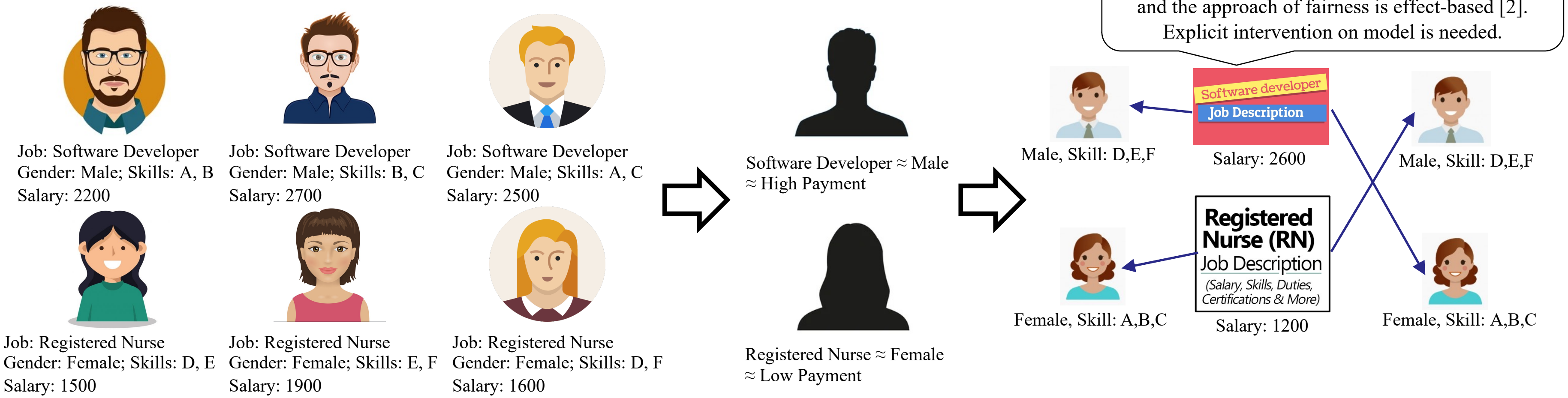


Interview opportunities are limited, which candidate(s) should get an interview opportunity?

Why Fairness in RecSys? Data Could be Biased

- Most RecSys models are ML models trained on some training data
  - Training data may encode social bias
  - Recommendation models may learn "shotcuts" for decision making
  - Model may echo or even reinforce the bias in training data

Just data debias is not enough because AI doesn't know which are sensitive features (e.g., gender) and the approach of fairness is effect-based [2]. Explicit intervention on model is needed.



Training Data

Strong correlationn between Job, Gender and Salary Level, while the skill feature shows less consistency among samples

Model

Model learns this strong correlation

Recommendation

Model echos/reinforces such correlation, the influence of skills is weakened by the strong data bias.

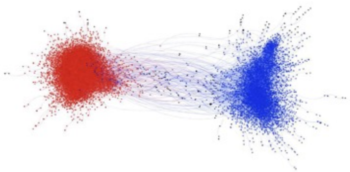
Potential Consequences of Unfairness in RecSys

Information Asymmetry

Knowing a piece of valuable information (e.g., a job opportunity) could change one's life

Matthew Effect

Advantaged users, items, or groups get further propagated by recommendations, sometimes not because of their good quality but because the recommendation model is dominated by their data

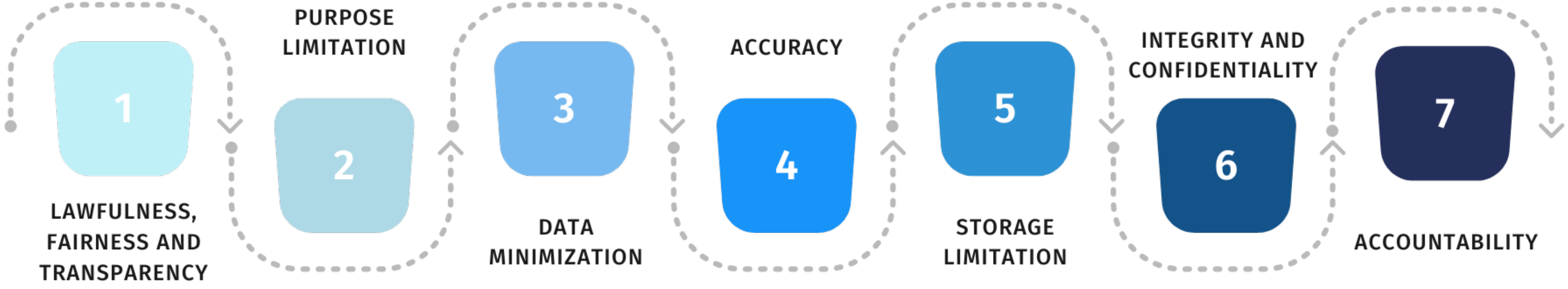


Echo Chambers

Unfair, undiversified exposure of news, messages, tweets, etc. may create echo chamber. Makes it difficult to explore new ideas and opinions different from one's own. Makes people feel like the whole world thinks the same way as they think. May even reinforce someone's extremist ideas

Fairness in RecSys: an AI Ethics Perspective

- Recommender systems as responsible AI
  - Provide fair decisions for users, item providers, and platform



7 Principles of EU GDPR Regulation

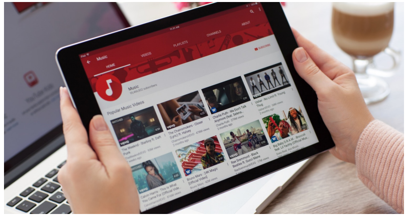
- Fairness often appears together with other responsible AI perspectives
  - e.g., transparency/explainability (honesty) of algorithmic decisions is the foundation of fairness

Fairness in RecSys: Beyond Ethics, a Utilitarian Perspective

- RecSys platforms should consider fairness for the sake of themselves
  - Not only for legal regulations, but for the sustainable/long-term development of the platform

An e-commerce example

Big retailors vs. Small retailors



A social network example

Star accounts vs. Grassroot accounts

If products from small retailors (e.g., family workshops) do not have fair exposure opportunity by e-commerce recommender system, they may eventually leave since they cannot survive in the platform, making the platform unsustainable.

Videos from famous accounts (e.g., a film star) usually get more attention, but if videos created by grassroot accounts do not have any exposure opportunity to users, they may leave the platform, making the platform's contents less diversified and even boring.

What exactly is Fairness in RecSys?

Many different perspectives:

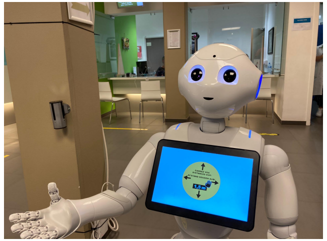
- Group Fairness vs. Individual Fairness
- User Fairness vs. Item Fairness
- Associative Fairness vs. Causal Fairness
- Single-sided Fairness vs. Multi-sided Fairness
- Static Fairness vs. Dynamic Fairness
- Short-term Fairness vs. Long-term Fairness
- Populational Fairness vs. Personalized Fairness

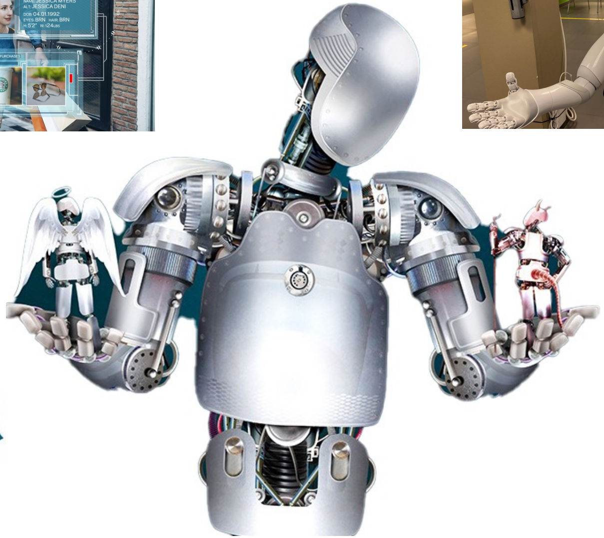
Outline

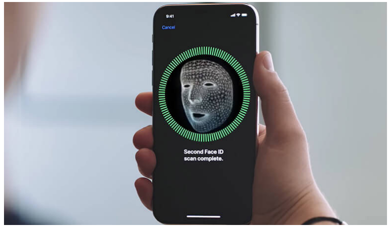
- Introduction:
  - Social Impact of Recommender System and Fairness
  - Motivation of Fairness in Recommender Systems
  - Relationship with AI Ethics
  - Beyond Ethics: a Utilitarian Perspective
- Fairness in Machine Learning:
  - Fairness in Classification
  - Fairness in Ranking
- Fairness in Recommendation:
  - Introduction
  - Taxonomy
  - Dataset and Evaluation
  - Challenge and Opportunity

Fairness in Machine Learning — Motivations

- Fairness matters because it has impact on everyone’s benefit.







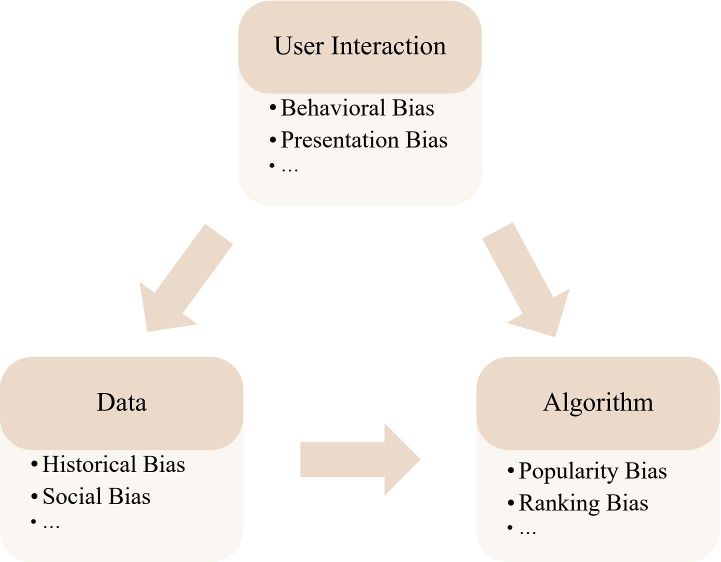
Fairness in Machine Learning — Causes

Data Bias

- Statistical Bias: non-random sample; record error
- Historical Bias: biased decision
- …

Algorithmic Bias

- Ranking Bias: exposure allocation
- Evaluation Bias: inappropriate benchmarks
- …



Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *arXiv preprint arXiv:1908.09635* (2019).
Castelnovo, Alessandro, et al. "The zoo of Fairness metrics in Machine Learning." *arXiv preprint arXiv:2106.00467* (2021).

Fairness in Machine Learning — Definitions

Individual Fairness	Counterfactual fairness
Group Fairness	Statistical parity $P(\hat{Y}\mid Z=0)=P(\hat{Y}\mid Z=1)$
Subgroup Fairness	Fairness holds over a large collection of subgroups defined by a class of functions

Fairness in Machine Learning — Methods

Pre-processing	In-processing	Post-processing
Try to transform the data so that the underlying discrimination is removed.	Try to modify the learning algorithms to remove discrimination during the model training process.	Perform after training by accessing a holdout set which was not involved during the training of the model.

Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *arXiv preprint arXiv:1908.09635* (2019)

Fairness in Machine Learning — Evaluation

The evaluation usually depends on the requirement of fairness.

- **Disparate Impact**: $P(\hat{y}=1|z=0)=P(\hat{y}=1|z=1)$
  - Evaluation: $DI=\left|P(\hat{y}=1|z=0)-P(\hat{y}=1|z=1)\right|$
- **False Positive Rate**: $P(\hat{y}\neq y|y=-1,z=0)=P(\hat{y}\neq y|y=-1,z=1)$
  - Evaluation: $DM_{FPR}=P(\hat{y}\neq y|z=0,y=-1)-P(\hat{y}\neq y|z=1,y=-1)$
- **False Negative Rate**: $P(\hat{y}\neq y|y=1,z=0)=P(\hat{y}\neq y|y=1,z=1)$
  - Evaluation: $DM_{FNR}=P(\hat{y}\neq y|z=0,y=1)-P(\hat{y}\neq y|z=1,y=1)$

Fairness in Machine Learning — Basic tasks

Fairness in Classification

Fairness in Ranking

Fairness in Classification — Introduction

Objective: Avoid unethical interference of protected attributes into the decision-making process.

Binary Classification: Fairness metrics can be expressed by **rate constraints** to regularize the classifier's positive or negative rates over different protected groups.

- Statistical parity:

  $P(\hat{Y}=1|Z=0)=P(\hat{Y}=1|Z=1)$

- Equality of Opportunity:

  $P(\hat{Y}=1|Z=0,Y=1)=P(\hat{Y}=1|Z=1,Y=1)$

…

RUTGERS

Fairness in Classification — Method

Pre-processing: [3][4][5][6]…

Pros:

The transformed dataset can be used to train any downstream algorithm.

Cons:

Unpredictable loss in accuracy;

May not remove unfairness on the test data.

In-processing: [7][8][9][10]…

Pros:

Good performance;

May higher flexibility for the trade-off.

Cons:

A non-convex optimization problem and not guarantee optimality.

Post-processing: [11][12][13]…

Pros:

No need to modify classifier;

Relatively good performance especially fairness measures.

Cons:

Cannot be used in cases where sensitive feature information is unavailable.

RUTGERS

Fairness in Classification — Dataset

Dataset	Size	Aera	Reference
UCI adult dataset	48,842 income records	Social	[14]
German credit dataset	1,000 credit records	Financial	[15]
Pilot parliaments benchmark	1,270 images	Facial images	[16]
WinoBias	3,160 sentences	Coreference resolution	[17]
Communities and crime	1,994 crime records	Social	[18]
COMPAS dataset	18,610 crime records	Social	[19]
Recidivism in juvenile justice	4,753 crime records	Social	[20]
Diversity in faces dataset	1 million images	Facial images	[21]

Mehrabi, Ninareh, et al. "A survey on bias and fairness in machine learning." *arXiv preprint arXiv:1908.09635* (2019).

Fairness in Classification

- **Fairness Concerns**: Introduce a flexible constraint-based framework to enable the design of fair margin-based classifiers.
- **Fairness Definitions**:
  - No disparate treatment: $P(\hat{y}|\boldsymbol{x}, z) = P(\hat{y}|\boldsymbol{x})$
  - No disparate impact: $P(\hat{y} = 1|z = 0) = P(\hat{y} = 1|z = 1)$
  - No disparate mistreatment:
    - False positive rate: $P(\hat{y} \neq y|y = -1, z = 0) = P(\hat{y} \neq y|y = -1, z = 1)$
    - False negative rate: $P(\hat{y} \neq y|y = 1, z = 0) = P(\hat{y} \neq y|y = 1, z = 1)$
    - …

“Zafar, Muhammad Bilal, et al. "Fairness Constraints: A Flexible Approach for Fair Classification." *J. Mach. Learn. Res.* 20.75 (2019): 1-42.

Fairness in Classification

- **Method:**

$$\begin{array}{ll} \text{minimize} & L(\boldsymbol{\theta}) \\ \text{subject to} & P(.|z=0) = P(.|z=1) \end{array} \begin{array}{l} \} \text{ Classifier loss function} \\ \} \text{ Fairness constraints} \end{array}$$

- No disparate impact: $P(\hat{y}=1|z=0) = P(\hat{y}=1|z=1)$

$$\mathrm{Cov}_{DI}(z, d_{\boldsymbol{\theta}}(\boldsymbol{x})) = \mathbb{E}[(z-\bar{z})d_{\boldsymbol{\theta}}(\boldsymbol{x})] - \mathbb{E}[(z-\bar{z})]\bar{d}_{\boldsymbol{\theta}}(\boldsymbol{x}) \approx \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x})$$

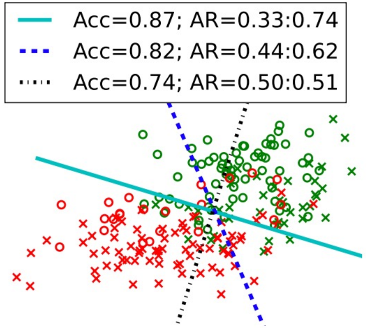
- Objective function for no disparate impact:

$$\begin{array}{ll} \text{minimize} & L(\boldsymbol{\theta}) \\ \text{subject to} & \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x}) \le c \\ & \frac{1}{N}\sum_{(\boldsymbol{x},z)\in\mathcal{D}}(z-\bar{z})\,d_{\boldsymbol{\theta}}(\boldsymbol{x}) \ge -c \end{array}$$

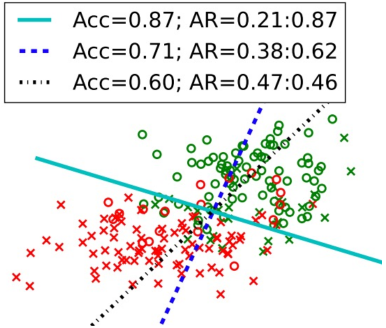
“Zafar, Muhammad Bilal, et al. "Fairness Constraints: A Flexible Approach for Fair Classification." *J. Mach. Learn. Res.* 20.75 (2019): 1-42

Fairness in Classification

- Simulate **disparate impact** in classification outcomes.
- Generate **two synthetic datasets** with different levels of correlation between a sensitive attribute and class labels.
- Train **logistic regression** classifiers on both the datasets.



$\phi = \pi/4$



$\phi = \pi/8$

"Zafar, Muhammad Bilal, et al. "Fairness Constraints: A Flexible Approach for Fair Classification." *J. Mach. Learn. Res.* 20.75 (2019): 1-42

Fairness in Classification

- Toshihiro Kamishima, et al. **"Fairness-aware classifier with prejudice remover regularizer."** In Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2012
- Aditya Krishna Menon, et al. **"The cost of fairness in binary classification."** FAT 2018
- Emmanouil Krasanakis, et al. **"Adaptive Sensitive Reweighting to Mitigate Bias in Fairness-aware Classification".** WWW 2018.
- Berk Ustun, et al. **"Fairness without Harm: Decoupled Classifiers with Preference Guarantees."** ICML 2019
- Muhammad Bilal Zafar et al. **"Fairness constraints: Mechanisms for fair classification."** arXiv:1507.05259 (2015).
- Yongkai Wu, et al. **"Fairness-aware Classification: Criterion, Convexity, and Bounds."** arXiv:cs.LG/1809.04737 (2018).
- Lingxiao Huang and Nisheeth Vishnoi. **"Stable and Fair Classification. "** ICML 2019
- Toon Calders and Sicco Verwer. **"Three naive Bayes approaches for discrimination-free classification."** Data Mining and Knowledge Discovery 21, 2 (2010), 277–292.

Fairness in Ranking — Introduction

List-wise definitions for fairness: depend on the entire list of results for a given query

Unsupervised criteria: the average **exposure** near the top of the ranked list to be **equal for different groups** [71][72][75]

Supervised criteria: the average **exposure** for a group to be proportional to the average **relevance** of that group's results to the query [65][67]

Fairness in Ranking — Method

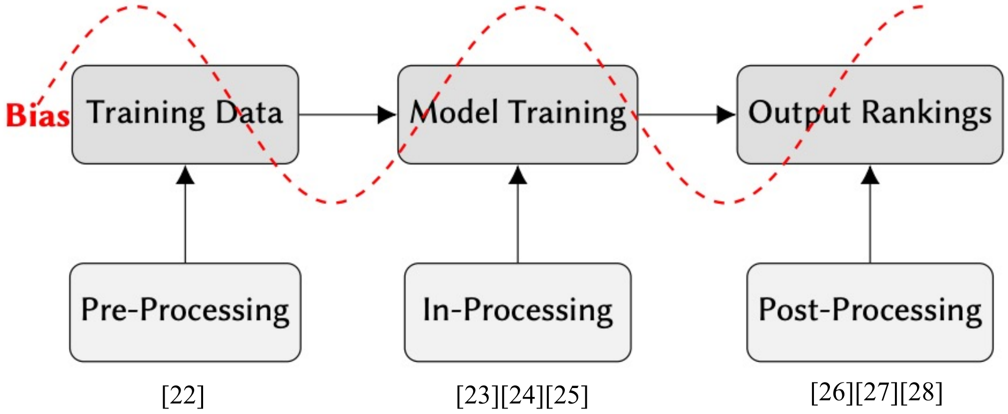


Figure Source: Zehlike, Meike, Ke Yang, and Julia Stoyanovich. "Fairness in Ranking: A Survey." *arXiv preprint arXiv:2103.14000* (2021).

Fairness in Ranking — Dataset

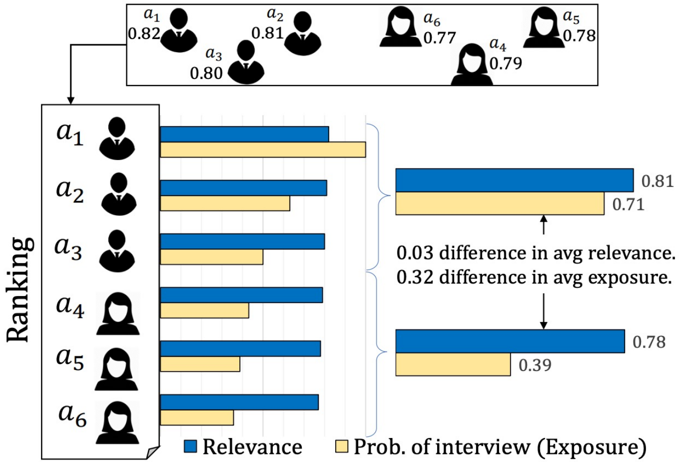
Dataset	Size	Sensitive Features	Reference
AirBnB	10,201 houses	gender of host	[29]
COMPAS	7,214 people	gender, race	[30]
CS departments	51 departments	department size, geographic region	[31]
Engineering students	5 queries, 650 students per query	gender, high school type	[32]
SAT	1.6M students	gender	[33]
German credit	1,000 people	gender, age	[34]
Forbes richest U.S.	400 people	gender	[35]
XING	40 candidates	gender	[36]

Zehlike, Meike, Ke Yang, and Julia Stoyanovich. "Fairness in Ranking: A Survey." *arXiv preprint arXiv:2103.14000* (2021).

Fairness in Ranking

- **Fairness Concerns**: A conceptual and computational framework that allows the formulation of fairness constraints on rankings in terms of **exposure allocation.**

- Job seeker example: a small difference in **relevance** can lead to a large difference in **exposure** (an opportunity) for the group of females.



Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*'2018.

Fairness in Ranking

- **Method:** $r \quad = \quad \text{argmax}_r \, \mathrm{U}(r|q)$ s.t. r is fair

- **Exposure** for a document d_i under a probabilistic ranking P as:

$$\text{Exposure}(d_i|\mathbf{P}) = \sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j \qquad \text{Exposure}(G_k|\mathbf{P}) = \frac{1}{|G_k|}\sum_{d_i \in G_k} \text{Exposure}(d_i|\mathbf{P})$$

- **Demographic Parity Constraints**:

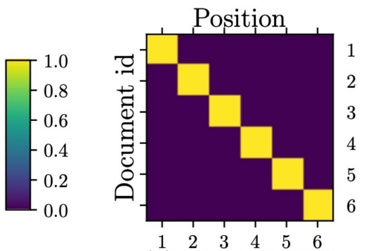
$$\text{Exposure}(G_0|\mathbf{P}) = \text{Exposure}(G_1|\mathbf{P}) \Leftrightarrow \mathbf{f}^T P\mathbf{v} = 0$$

$$(\text{with } \mathbf{f}_i = \frac{\mathbb{1}_{d_i \in G_0}}{|G_0|} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|})$$

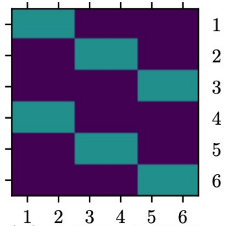
Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD'*2018

Fairness in Ranking

- Figure (a) is optimal unfair ranking that maximizes DCG.
- Figure (b) is optimal fair ranking under demographic parity.
- Compared to the DCG of the unfair ranking, the optimal fair ranking has slightly **lower utility** with a DCG.



(a) DCG=3.8193



(b) DCG=3.8031

Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*'2018

Fairness in Ranking

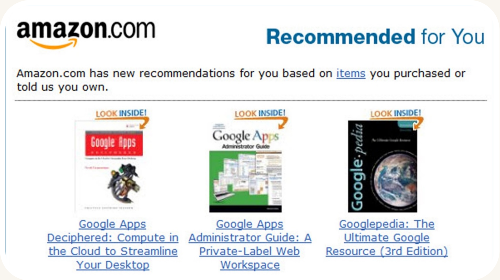
- Biega, Asia J.,et al. **"Equity of attention: Amortizing individual fairness in rankings."** *SIGIR* 2018.
- Zehlike, Meike, et al. "Fa*ir: A fair top-k ranking algorithm." *CIKM* 2017.
- Singh, Ashudeep, and Thorsten Joachims. **"Policy learning for fairness in ranking."** *arXiv preprint arXiv:1902.04056* (2019).
- Yadav, Himank, et al. **"Fair learning-to-rank from implicit feedback."** *arXiv preprint arXiv:1911.08054* (2019).
- Yang, Ke, and Julia Stoyanovich. **"Measuring fairness in ranked outputs."** *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. 2017.
- Narasimhan, Harikrishna, et al. **"Pairwise fairness for ranking and regression."** *AAAI* 2020.
- Celis, L. Elisa, et al. **"Ranking with fairness constraints."** *arXiv preprint arXiv:1704.06840* (2017).
- Zehlike, Meike, and Carlos Castillo. **"Reducing disparate exposure in ranking: A learning to rank approach."** *Proceedings of The Web Conference 2020.*
- Vogel, Robin, et al. **"Learning Fair Scoring Functions: Fairness Definitions, Algorithms and Generalization Bounds for Bipartite Ranking."** *arXiv preprint arXiv:2002.08159* (2020).

Outline

- Introduction:
  - Social Impact of Recommender System and Fairness
  - Motivation of Fairness in Recommender Systems
  - Relationship with AI Ethics
  - Beyond Ethics: a Utilitarian Perspective
- Fairness in Machine Learning:
  - Fairness in Classification
  - Fairness in Ranking
- Fairness in Recommendation:
  - Introduction
  - Taxonomy
  - Dataset and Evaluation
  - Challenge and Opportunity

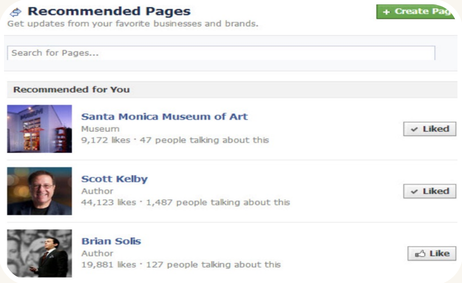
Fairness in Recommendation — Motivation

- Recommender systems are gaining critical impacts on human decision making.



Shopping

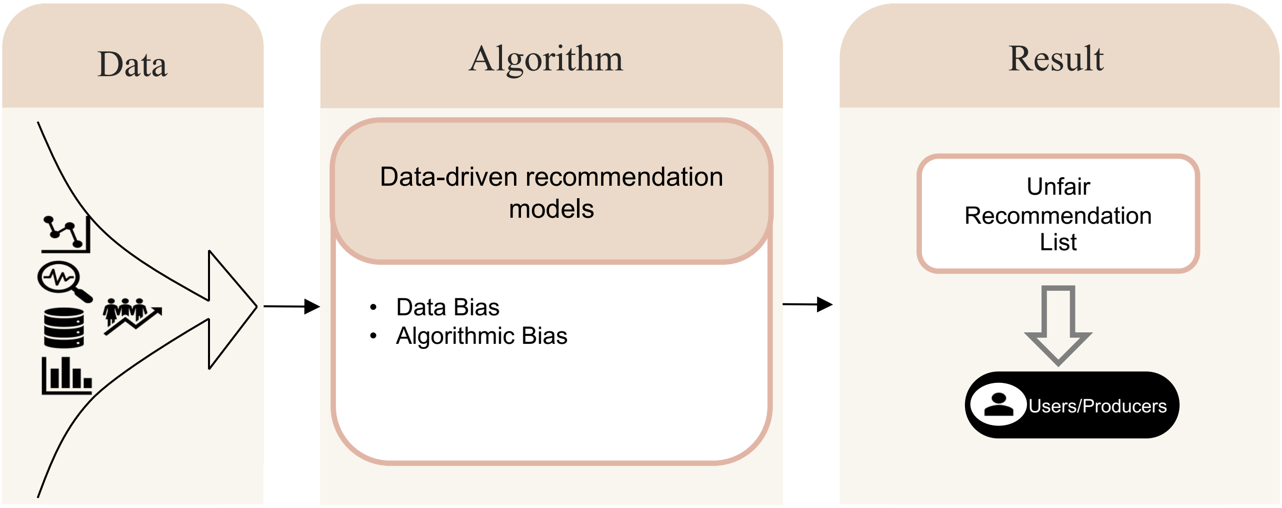
Entertainment



Social Media

Fairness in Recommendation — Motivation

- It is crucial to address the potential unfairness problems in recommendations.



Fairness in Recommendation — Challenges

More Perspectives

Multiple Models And Goals

Extreme Data Sparsity

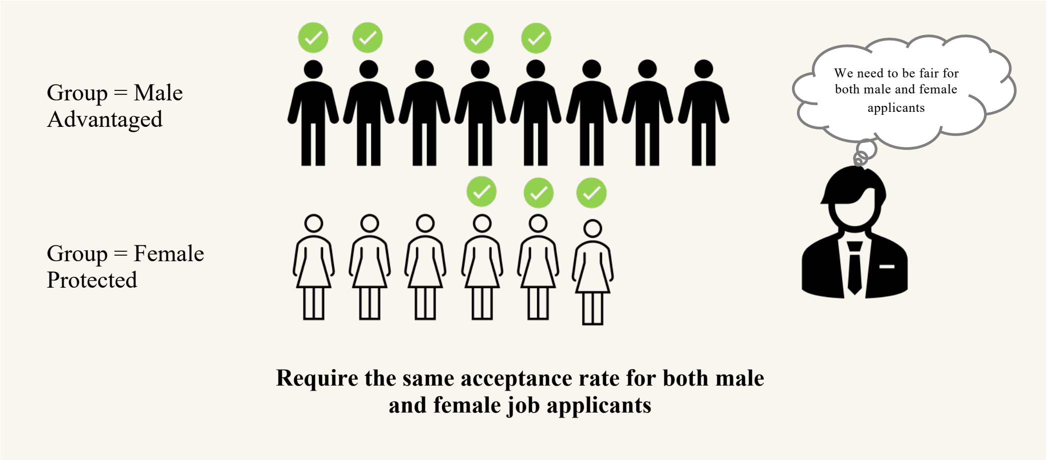
Dynamics

Taxonomies

- Group vs. Individual
- User vs. Item
- Association vs. Causality
- Single-sided vs. Multi-sided
- Static vs. Dynamic

Group Fairness vs. Individual Fairness

Group fairness requires that the protected groups should be treated similarly to the advantaged group.



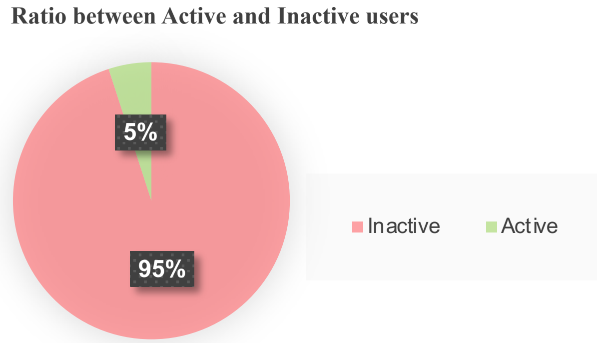
Group Fairness vs. **Individual Fairness**

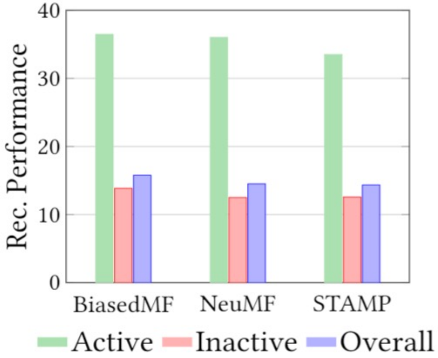
- Individual fairness requires that the similar individual should be treated similarly.

Image source: https://mitibmwatsonailab.mit.edu/wp-content/uploads/2020/04/Leadspace-GettyImages-598952582.jpg

Group Fairness in Recommendation

- **Fairness concerns**: The **unfair recommendation quality** between **user groups** with different activity levels, e.g., number of interactions.
- Unfairness of current recommender systems:
  - Active users only account for a **small** proportion of users.
  - The average recommendation quality on the small group (*active*) is **significantly better** than that on the remaining majority of users (*inactive*) for all baselines.





Li.Y et al. "User-oriented Fairness in Recommendation" WWW'21.

RUTGERS

Group Fairness in Recommendation

Fairness-aware Algorithm: A re-ranking method with user-oriented group fairness constrained on the recommendation lists generated from any base recommender algorithm.

$$\max_{\mathbf{W}_{ij}} \sum_{i=1}^{n}\sum_{j=1}^{N} \mathbf{W}_{ij}\boxed{S_{i,j}}$$

Preference of user i in terms of item j

$$\text{s.t.} \quad \boxed{\mathrm{UGF}\left(Z_1, Z_2, \mathbf{W}\right) < \varepsilon}$$

Fairness constraint

$$\boxed{\sum_{j=1}^{N} \mathbf{W}_{ij} = K, \mathbf{W}_{ij} \in \{0, 1\}}$$

Top-K list

Experiment Results: Improve fairness; Improve recommendation quality of overall and disadvantaged users. However, the performance of advantaged users is reduced to satisfy our fairness requirement.

			Beauty			
			Overall	Adv.	Disadv.	UGF
BiasedMF	F1	Orig.	14.27	30.68	12.77	17.91
		Fair	**15.06**	19.18	**14.68**	**4.50**
	NDCG	Orig.	43.25	67.79	41.00	26.79
		Fair	**43.97**	52.51	**43.19**	**9.32**

Improvement of overall accuracy

Disadv. ⇧

Adv. ⇩

Improvement of fairness

Li.Y et al. "User-oriented Fairness in Recommendation" WWW'21.

Group Fairness in Recommendation

Other works:

- Fu et al. [37] require to impair the group unfairness problem in the context of explainable recommendation over knowledge graphs with a fairness constrained approach.

- Both [38] and [39] categorize different types of multi-stakeholder platforms and the different group fairness properties they desired.

Individual Fairness in Recommendation

- **Fairness concerns**: the position bias which leads to disproportionately less attention being paid to low-ranked subjects.
- No single ranking can achieve individual attention fairness.
- **Equity of Amortized Attention**: A sequence of rankings $\{1,2,\dots m\}$ offer equity of amortized attention if each subject u receives cumulative attention proportional to her cumulative relevance:

attention

$$\frac{\boxed{\sum_{l=1}^{m} a_{i1}^{l}}}{\boxed{\sum_{l=1}^{m} r_{i1}^{l}}} = \frac{\sum_{l=1}^{m} a_{i2}^{l}}{\sum_{l=1}^{m} r_{i2}^{l}}, \ \forall u_{i1}, u_{i2}$$

relevance

Biega, A. J. et al. “Equity of Attention: Amortizing Individual Fairness in Rankings” SIGIR’18.

Individual Fairness in Recommendation

- **Method (Offline optimization):**

$$\text{minimize} \quad \boxed{\sum_i |A_i - R_i|} \quad \text{Fairness}$$

$$\text{subject to} \quad \boxed{NDCG\text{-}quality@k(\rho^j, \rho^{j*}) \geq \theta, \; j = 1, \ldots, m.} \quad \text{Ranking quality}$$

- **Experiment Results:**
  - **Improving equity of attention is crucial**: the discrepancy between the attention received and the deserved attention can be substantial.
  - Improving equity of attention can often be done **without sacrificing much quality** in the rankings.

Biega, A. J. et al. "Equity of Attention: Amortizing Individual Fairness in Rankings" SIGIR'18.

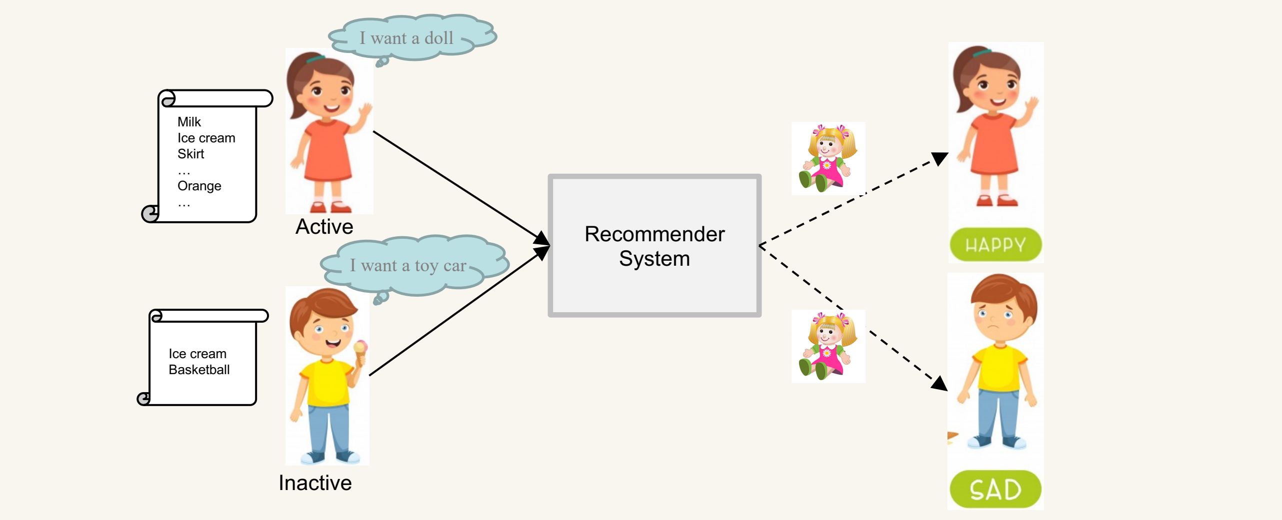
Individual Fairness in Recommendation

Other Works:

- Patro et al. [40] view individual fairness from both producers and customers sides, and response to the question of the long-term sustainability of two-sided platforms.
- Li et al. [80] consider personalized fairness for users in recommendations, i.e., users' personalized demands for fairness. For example, some users may care more about gender, while others care more about age.

User Fairness vs. Item Fairness

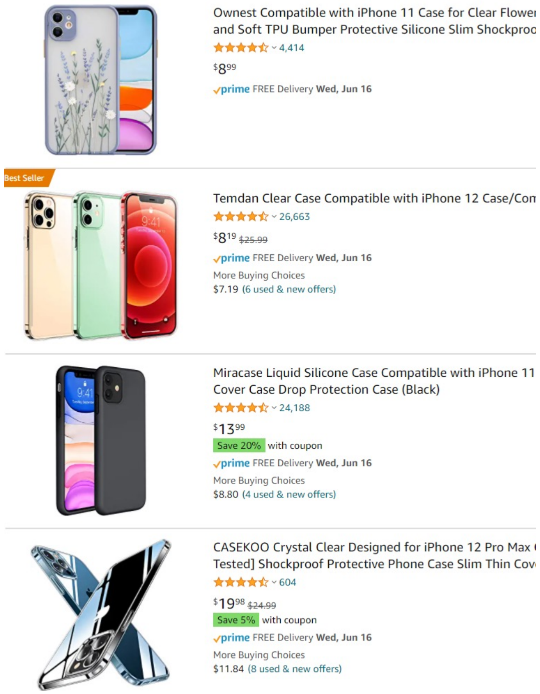
Fairness on user side: Fairness requirements in recommender systems may come from users.



User Fairness vs. **Item Fairness**

- Fairness on Item side: Fairness requirements in recommender systems may come from items (Products/Producers).

- For example, we search for “phone case” but the system ranks accessories for iPhone on top but quite few for other brands, which is an item-side unfairness.



User Fairness in Recommendation

- **Group Recommendation**: recommend items to groups of users whose preferences can be different from each other.
- **Fairness Concerns**: maximize the satisfaction of each group member while minimizing the unfairness (the imbalance of user utilities inside the group) between them.
- **Fairness Definitions**:
  - Least Misery: $F_{LM}(g, I) = \min\{\boxed{U(u, I)}, \forall u \in g\}$
  - Variance: $F_{Var}(g, I) = 1 - Var(\{U(u, I), \forall u \in g\})$
  - Jain's Fairness: $F_J(g, I) = \frac{(\sum_{u \in g} U(u, I))^2}{|U| \cdot \sum_{u \in g} U(u, I)^2}$
  - Min-Max Ratio: $F_M(g, I) = \frac{\min\{U(u, I), \forall u \in g\}}{\max\{U(u, I), \forall u \in g\}}$

The individual utility of user *u* in group *g* when a set of items *I* are recommended to the group.

Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys'17

User Fairness in Recommendation

- **Method**:
  - The **Social Welfare** ($SW(g,I)$): overall utility of all users inside the group g given a group recommendation I.
  - The **Fairness** ($F(g,I)$): a function of $U(u,I), \forall u \in g, \forall I$.
  - Multi-Objective Optimization: $\lambda \cdot SW(g,I) + (1-\lambda) \cdot F(g,I)$
- **Experiment Results**: The results indicate that considering fairness can improve the quality of group recommendation.

λ, RG	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
F@K	0.0260	0.0817	0.0877	0.0953	0.1019	0.1041	0.1046	0.1053	0.1058	0.1062	0.1062
NDCG@K	0.0697	0.2200	0.2287	0.2334	0.2394	0.2423	0.2440	0.2421	0.2459	0.2478	0.2476

Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys'17

User Fairness in Recommendation

Other Works:

- Leonhardt et al. [41] quantify the user unfairness caused by the post-processing algorithms which have the original goal of improving diversity in recommendations.

- Abdollahpouri et al. [42] see the problem from the users' perspective with finding how popularity bias causes the recommendations to deviate from what the user expects to get from the recommender system.

Item Fairness in Recommendation

- **Fairness Concerns:** focus on the risk to groups of items from being under-recommended

Monotonic ranking from predictions

- Pairwise Accuracy:

$$P(\boxed{g(f_\theta(\mathbf{q},\mathbf{v}_j)) > g(f_\theta(\mathbf{q},\mathbf{v}_{j'}))} | \boxed{y_{\mathbf{q},j} > y_{\mathbf{q},j'}}, j, j' \in \mathcal{R}_\mathbf{q}) \quad c_\mathbf{q}(j,j') \triangleq \mathbb{1}[g(f_\theta(\mathbf{q},\mathbf{v}_j)) > g(f_\theta(\mathbf{q},\mathbf{v}_{j'}))]$$

User click feedback

- Pairwise Fairness:

$$P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, \boxed{s_j = 0,} z_{\mathbf{q},j} = \tilde{z}) = P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, s_j = 1, z_{\mathbf{q},j} = \tilde{z}), \quad \forall \tilde{z}$$

Group label

- Inter-Group Pairwise Fairness:

$$P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, \boxed{s_j = 0, s_{j'} = 1,} z_{\mathbf{q},j} = \tilde{z}) = P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, s_j = 1, s_{j'} = 0, z_{\mathbf{q},j} = \tilde{z}), \forall \tilde{z}$$

- Intra-Group Pairwise Fairness:

$$P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, \boxed{s_j = s_{j'} = 0,} z_{\mathbf{q},j} = \tilde{z}) = P(c_\mathbf{q}(j,j') | y_{\mathbf{q},j} > y_{\mathbf{q},j'}, s_j = s_{j'} = 1, z_{\mathbf{q},j} = \tilde{z}), \forall \tilde{z}$$

Beutel, Alex, et al. “Fairness in Recommendation Ranking through Pairwise Comparisons” SIGKDD’19

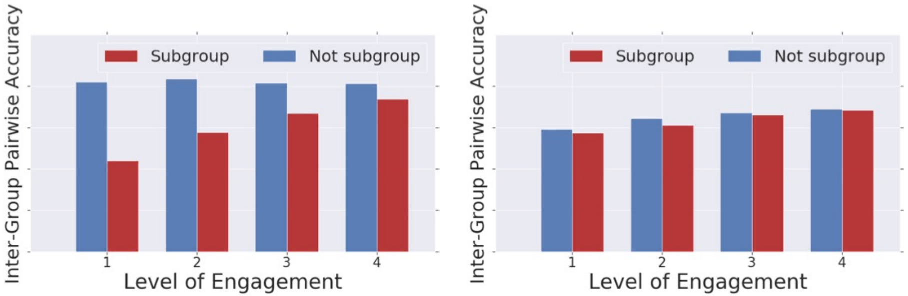
Item Fairness in Recommendation

- **Method:**

$$\min_{\theta} \overbrace{\left(\sum_{(\mathbf{q},j,y,z)\in\mathcal{D}} L(f_\theta(\mathbf{q},\mathbf{v}_j),(y,z)) \right)}^{\text{Recommender Loss}} + \overbrace{\left|\mathrm{Corr}_{\mathcal{P}}(A,B)\right|}^{\text{Fairness Penalty}}$$

$$A = (g(f_\theta(\mathbf{q},\mathbf{v}_j)) - g(f_\theta(\mathbf{q},\mathbf{v}_{j'})))(y - y')$$
$$B = (s_j - s_{j'})(y - y')$$

- **Experiment Results:**
  - The subgroup items are significantly under-ranked relative to the non-subgroup items.
  - The regularization effectively closes the gap in the inter-group pairwise fairness metric.



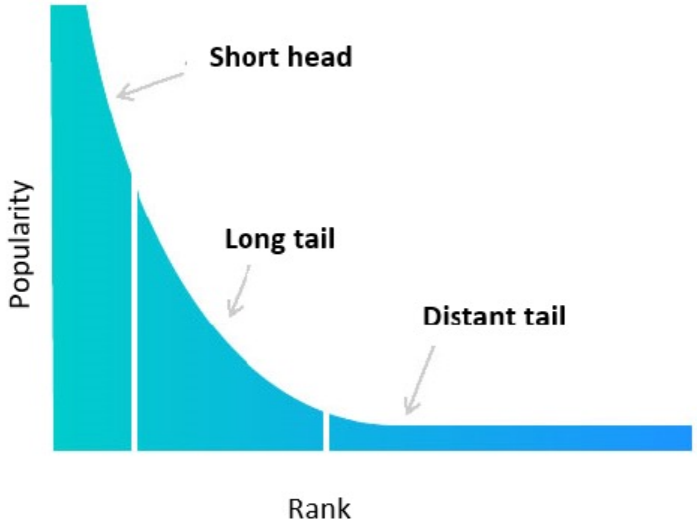
(a) Original **(b) After Pairwise Reg.**

Beutel, Alex, et al. "Fairness in Recommendation Ranking through Pairwise Comparisons" SIGKDD'19

Item Fairness in Recommendation

Other Works:

- Many works about the **popularity bias** problem in recommendations.
- Often solved by increasing the number of recommended unpopular items (long-tail items) or otherwise the overall catalog coverage in these researches [43-45].



Abdollahpouri, et al. "Managing popularity bias in recommender systems with personalized re-ranking." *arXiv preprint arXiv:1901.07555* (2019).

Associative Fairness vs. Causal Fairness

Find the **discrepancy of statistical metrics** between individuals or sub-populations.

In **binary classification**, fairness metrics can be represented by regularizing the classifier's positive or negative rates over different protected groups.

Associative Fairness vs. **Causal Fairness**

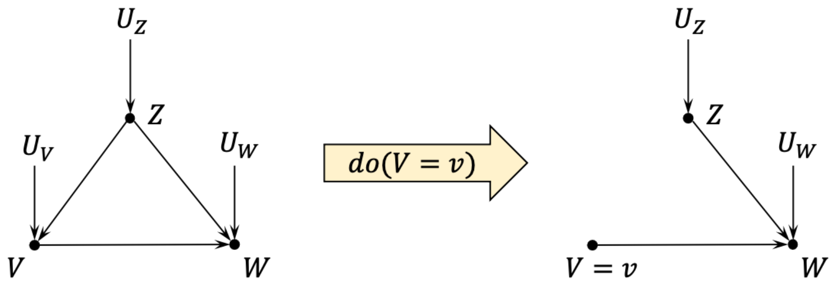
- Fairness cannot be well assessed only based on association notions [46-49].
- Difference:
  - Reason about the **causal relations** between the protected features and the model outcomes.
  - Leverage **prior knowledge** about the world structure in the form of causal models, help to understand the propagation of variable changes in the system.

Causal Fairness

- **Methods**:
  - Intervention
  - Counterfactual

- **Causal graph:** A directed acyclic graph which is used to capture the causal relations between variables, where nodes represent variables and directed edges represent a causal influence between the corresponding variables.

- A **“what if”** statement in which the “if” portion is **unreal** or **unrealized**, is known as a counterfactual.



RUTGERS

Causal Fairness

- **Disparate Impact**:
  - **Total Effect**: $TE_{a_1,a_0}(y) = P(y_{a_1}) - P(y_{a_0})$
  - **Effect of Treatment on the Treated**: $ETT_{a_1,a_0}(y) = P(y_{a_1} \mid a_0) - P(y \mid a_0)$
  - …
- **Disparate Treatment**:
  - **Direct Effect**: the causal effect along the causal path from the sensitive feature to the final decision
  - **Indirect Effect**: the causal effect along the causal path through proxy features
  - **Path-Specific Effect**: the causal effect over specific paths.

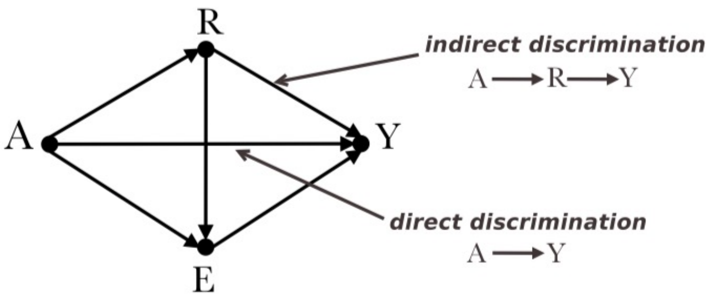
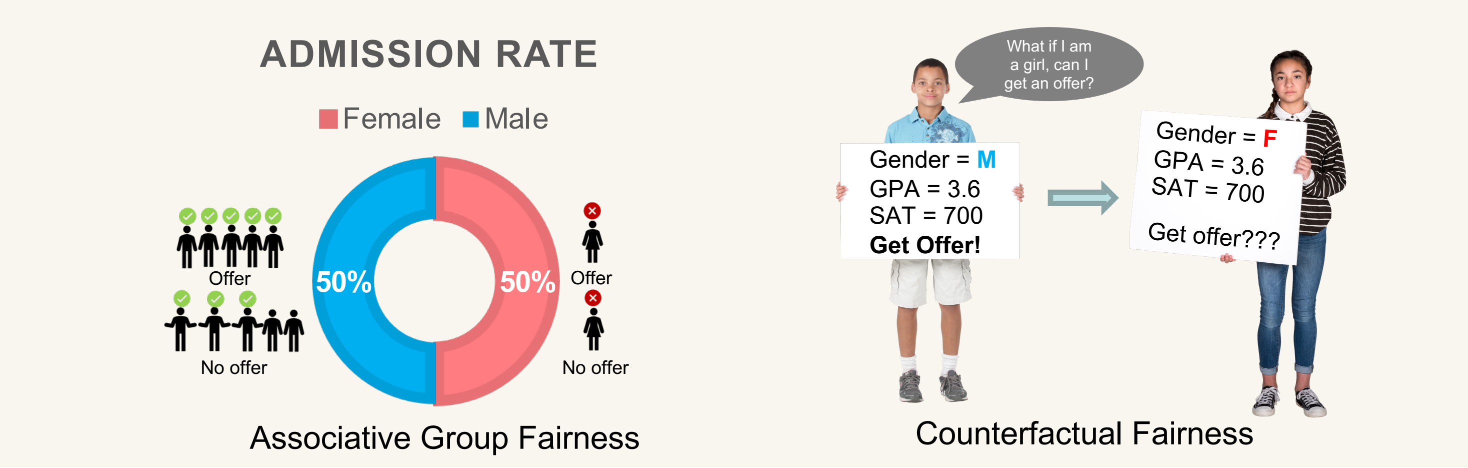


Figure Source: Makhlouf, Karima, et al. "Survey on Causal-based Machine Learning Fairness Notions." *arXiv preprint arXiv:2010.09553* (2020).

Counterfactual fairness

- Counterfactual fairness is an individual-level causal-based fairness notion. It requires that for any possible individual, the predicted result of the learning system should be the same in the **same** in the **counterfactual world** as in the **real world.**



Associative Fairness in Recommendation

- **Fairness Concerns**: study fairness in collaborative-filtering recommender systems; propose four new metrics that address different forms of unfairness.

Average predicted score from disadvantaged users

Average ratings from disadvantaged users

Average predicted score from advantaged users

Average ratings from advantaged users

- **Fairness Definitions**:
  - Value Fairness: $U_{\text{val}} = \frac{1}{n}\sum_{j=1}^{n}\left|\left(\boxed{\mathrm{E}_g[y]_j} - \boxed{\mathrm{E}_g[r]_j}\right) - \left(\boxed{\mathrm{E}_{\neg g}[y]_j} - \boxed{\mathrm{E}_{\neg g}[r]_j}\right)\right|$
  - Absolute Fairness: $U_{\text{abs}} = \frac{1}{n}\sum_{j=1}^{n}\left|\left|\mathrm{E}_g[y]_j - \mathrm{E}_g[r]_j\right| - \left|\mathrm{E}_{\neg g}[y]_j - \mathrm{E}_{\neg g}[r]_j\right|\right|$
  - Underestimation unfairness: $U_{\text{under}} = \frac{1}{n}\sum_{j=1}^{n}\left|\max\{0, \mathrm{E}_g[r]_j - \mathrm{E}_g[y]_j\} - \max\{0, \mathrm{E}_{\neg g}[r]_j - \mathrm{E}_{\neg g}[y]_j\}\right|$
  - Overestimation unfairness: $U_{\text{over}} = \frac{1}{n}\sum_{j=1}^{n}\left|\max\{0, \mathrm{E}_g[y]_j - \mathrm{E}_g[r]_j\} - \max\{0, \mathrm{E}_{\neg g}[y]_j - \mathrm{E}_{\neg g}[r]_j\}\right|$

Yao, Sirui, and Bert Huang. "Beyond Parity: Fairness Objectives for Collaborative Filtering" NIPS'17

Rutgers

Associative Fairness in Recommendation

- **Method**: $$\min_{\boldsymbol{P},\boldsymbol{Q},\boldsymbol{u},\boldsymbol{v}} \underbrace{\boxed{J(\boldsymbol{P},\boldsymbol{Q},\boldsymbol{u},\boldsymbol{v})}}_{\text{Loss for recommender model}} + \underbrace{\boxed{U}}_{\text{Fairness constraint}}$$

- **Experiment Results**: the experiments on synthetic and real data show that minimization of these forms of unfairness is possible with no significant increase in reconstruction error.

Unfairness	Error	Value	Absolute	Underestimation	Overestimation	Non-Parity
None	0.887 ± 1.9e-03	0.234 ± 6.3e-03	0.126 ± 1.7e-03	0.107 ± 1.6e-03	0.153 ± 3.9e-03	0.036 ± 1.3e-03
Value	0.886 ± 2.2e-03	**0.223 ± 6.9e-03**	0.128 ± 2.2e-03	**0.102 ± 1.9e-03**	**0.148 ± 4.9e-03**	0.041 ± 1.6e-03
Absolute	0.887 ± 2.0e-03	0.235 ± 6.2e-03	**0.124 ± 1.7e-03**	0.110 ± 1.8e-03	0.151 ± 4.2e-03	0.023 ± 2.7e-03
Under	0.888 ± 2.2e-03	0.233 ± 6.8e-03	0.128 ± 1.8e-03	**0.102 ± 1.7e-03**	0.156 ± 4.2e-03	0.058 ± 9.3e-04
Over	**0.885 ± 1.9e-03**	0.234 ± 5.8e-03	**0.125 ± 1.6e-03**	0.112 ± 1.9e-03	**0.148 ± 4.1e-03**	0.015 ± 2.0e-03
Non-Parity	0.887 ± 1.9e-03	0.236 ± 6.0e-03	0.126 ± 1.6e-03	0.110 ± 1.7e-03	0.152 ± 3.9e-03	**0.010 ± 1.5e-03**

Yao, Sirui, and Bert Huang. "Beyond Parity: Fairness Objectives for Collaborative Filtering" NIPS'17

Causal Fairness in Recommendation

- **Fairness Concerns**: Counterfactual fairness for users in recommendations.

- **Definition:** A recommender model is *counterfactually fair* if for any possible user u with features $X = x$ and $Z = z$, for all L, and for any value z' attainable by Z:

$$P(\boxed{L_z} \mid \boxed{X} = x, \boxed{Z} = z) = P(L_{z'} \mid X = x, Z = z)$$

Top-N recommendation list for user u with sensitive features z

Insensitive features

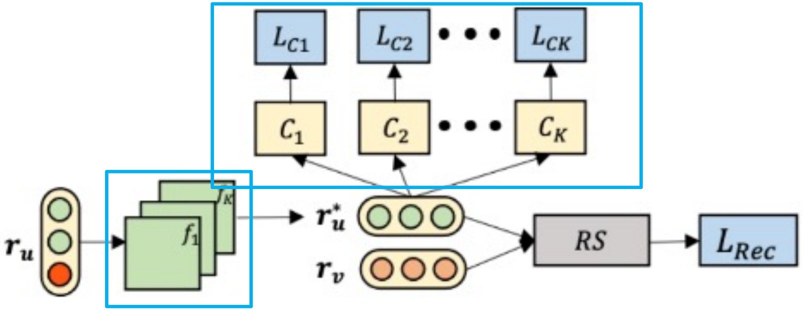
Sensitive features

Li. Y et al. "Towards Personalized Fairness based on Causal Notion" SIGIR'21

Causal Fairness in Recommendation

- **Method:** Generate feature independent user embeddings through *adversary learning*.
  - **Filter Module**: filter the information about sensitive features from user embeddings
  - **Discriminator module:** predict the sensitive features from the learned user embeddings.
- **Experiment Results**:
  - Improve fairness
  - A little sacrifice on recommendation performance

Discriminator module



Filter module

		MoiveLens		
		AUC-G	AUC-A	AUC-O
PMF	Orig.	0.7697	0.8428	0.6024
	SM	**0.5389**	**0.5560**	**0.5289**
	CM	0.5532	0.5951	0.5396

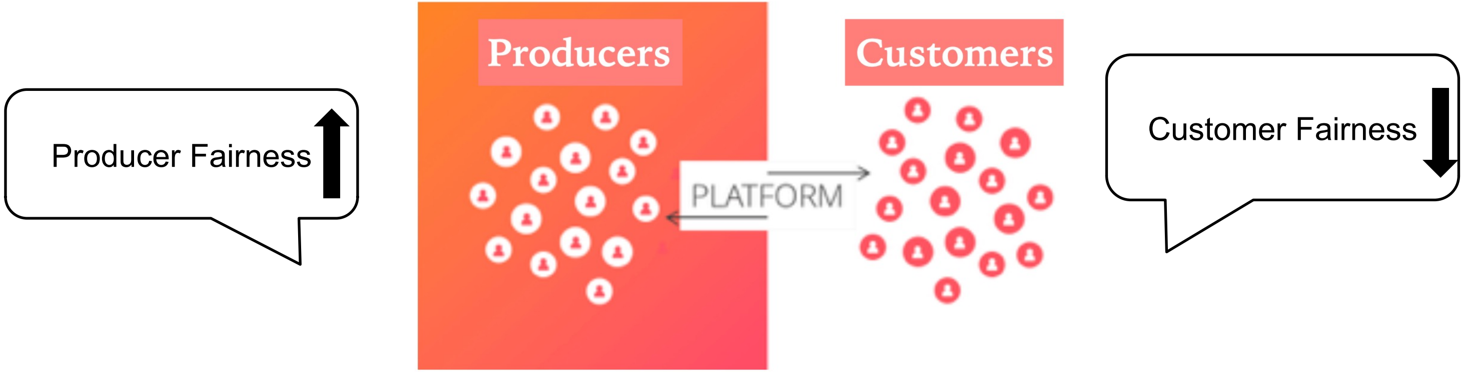
Li. Y et al. “Towards Personalized Fairness based on Causal Notion” SIGIR’21

Single-sided vs. Multi-sided Fairness

- Most research on the fairness of recommender systems is conducted either from the perspective of **customers** or from the perspective of product (or service) **providers**, which is also known as **single-sided** fairness.
- Fairness, that considers both **customer-side** fairness and **provider-side** fairness, is known as **multi-sided** fairness.

Multi-sided Fairness

- Why multi-sided fairness?
  - When fairness is guaranteed for one side, the fairness and rights of the other side might sacrifice [83,84]



Multi-sided Fairness

- How to approach multi-sided fairness?
  - Usually, the two-sided objective is a linear interpolation of consumer and producer fairness metrics [81,82,83].

$$\begin{array}{ll} \underset{M}{\text{minimize}} & \lambda \cdot \text{inequality}_D(M) + (1 - \lambda) \cdot \text{inequality}_C(M) \\ \text{subject to} & \text{constraints ensuring a correct matching.} \end{array}$$

Suhr et al. "Two-Sided Fairness for Repeated Matchings in Two-Sided Markets: A Case Study of a Ride-Hailing Platform" KDD'19

Static vs. Dynamic Fairness

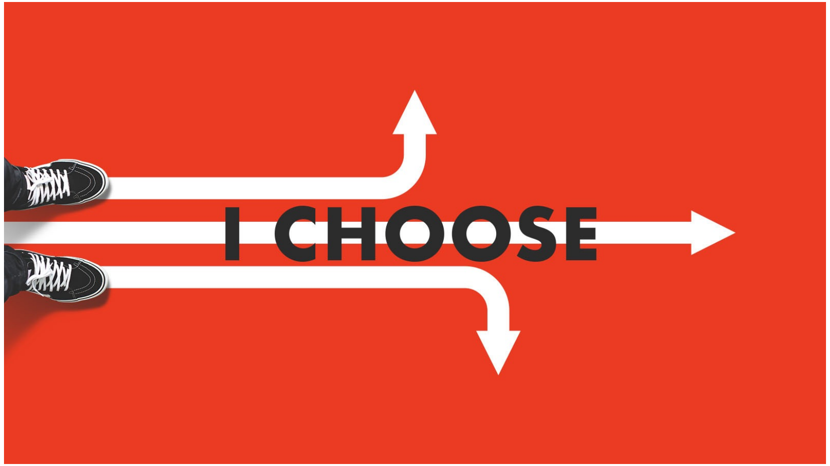
- **Static fairness** means the protected attribute or group labels (i.e., gender or race) are fixed throughout the entire ranking or recommendation process.
- **Dynamic fairness** considers the dynamic factors in the environment, such as the changes in item utility or attributes in recommendation environment, and learns a strategy to accommodate such dynamics.

Static vs. Dynamic Fairness

- Why consider dynamic fairness?
  - In real world, biases are usually dynamic rather than static. For example,
    - New items will come into the item pool;

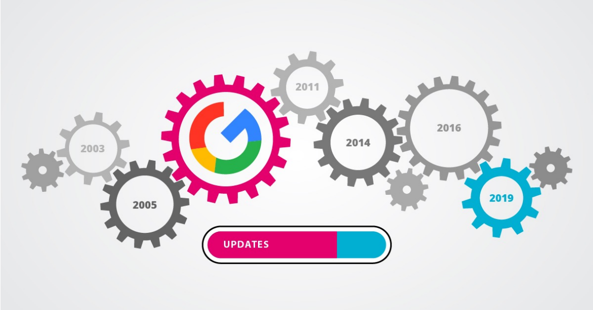
Static vs. Dynamic Fairness

- Why consider dynamic fairness?
  - In real world, biases are usually dynamic rather than static. For example,
    - Users experience many new items and may change their preferences;



Static vs. Dynamic Fairness

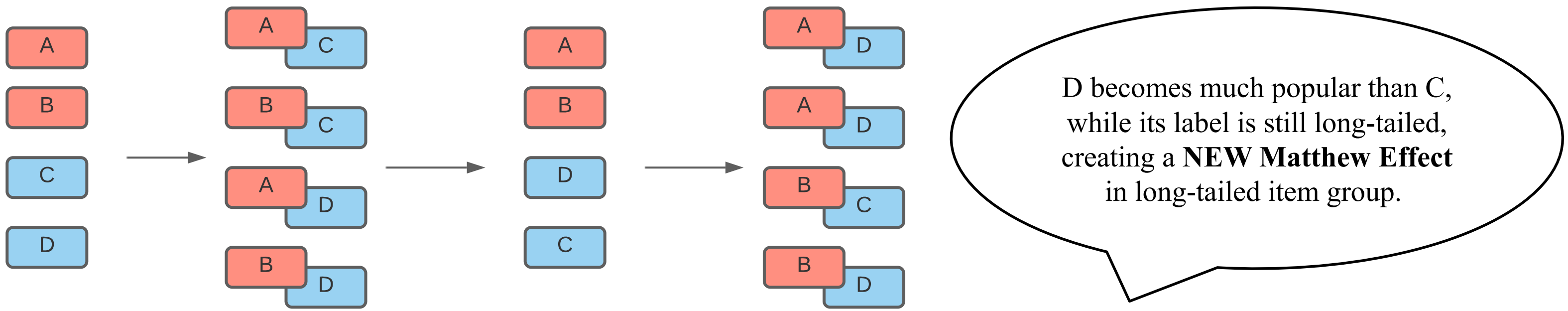
- Why consider dynamic fairness?
  - In real world, biases are usually dynamic rather than static. For example,
    - The recommendation system will update its recommendation strategy periodically.



Static vs. Dynamic Fairness

- How to achieve dynamic fairness?

Ge et al. [91] proposed to model the dynamic long-term fairness in recommendation with respect to **dynamically changing group labels** through a fairness-constrained reinforcement learning framework.



Ge et al. "Towards Long-term Fairness in Recommendation". WSDM'21

Static vs. Dynamic Fairness

The problem is formulated as **Constrained MDP** (Markov Decision Process).

State: state s_t of a user

H_t - user's most recent positive interaction history

Demographic information (if exists).

Action: a recommendation list a_t={$a_t{}^1$, … , $a_t{}^K$} with current state s_t.

Reward: the immediate feedback $R(st,at)$ given the action a_t and the user state s_t

Typical user feedback includes click, skip, or purchase, etc.

Cost: a cost value $C(s_t,a_t)$ given by the problem-specific cost function

i.e., #items that come from the sensitive group

Discount rate: γ_r and γ_c:

$\gamma_r \in [0,1]$ is for long-term rewards

$\gamma_c \in [0,1]$ is for long-term costs.

Ge et al. "Towards Long-term Fairness in Recommendation". WSDM'21

Static vs. Dynamic Fairness

- How to achieve dynamic fairness?

  Define Exact-K fairness,

$$\frac{\text{Exposure}_t\ (G_0)}{\text{Exposure}_t\ (G_1)} \leq \alpha$$

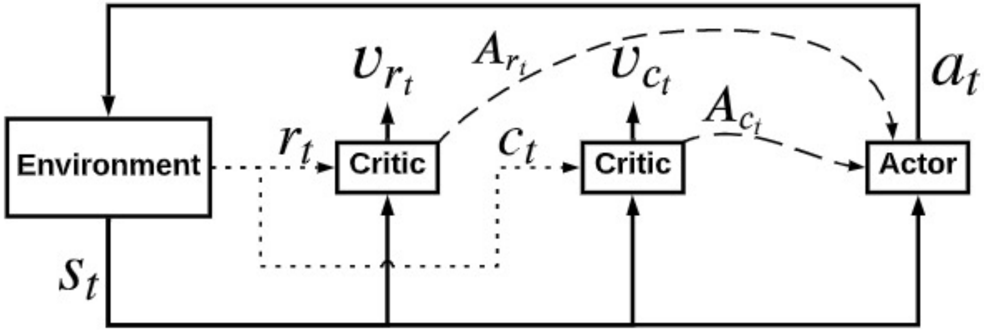
$$\text{Exposure}_t\ (G_0)\ \leq \alpha \text{Exposure}_t\ (G_1)$$

Ge et al. "Towards Long-term Fairness in Recommendation". WSDM'21

Static vs. Dynamic Fairness

- How to achieve dynamic fairness?

  Finally, they used Constraint Policy Optimization to solve the above problem.



Using two critics – a fairness critic and a utility critic to learn fairness and utility in a dynamic reinforcement learning framework.

Ge et al. "Towards Long-term Fairness in Recommendation". WSDM'21

Applications of Fairness-aware RecSys

- Ride-hailing (Uber, Lyft): Sühr et al. [81]
- Ecommerce (Amazon, Etsy): Patro et al. [82]
- Content streaming (Spotify, YouTube): Htun et al. [85]
- Social Media (Twitter, LinkedIn): Vasudevan et al. [86], Geyik et al. [87]
- Cyber-Physical Systems (e-Vehicle charging): Wang et al. [90]

Evaluation of Fairness

- Measuring User-side Fairness: Based on the general definition of user fairness, Fu et al. [89] defined GRU as a measurement.
- **Group Recommendation Unfairness (GRU)**

$$GRU(G_1, G_2, \mathbf{Q}) = \left| \frac{1}{|G_1|} \sum_{i \in G_1} \mathcal{F}(\mathbf{Q}_i) - \frac{1}{|G_2|} \sum_{i \in G_2} \mathcal{F}(\mathbf{Q}_i) \right|$$

F refers to a metric that scores the recommendation quality such that F (Qi) denotes the recommendation quality for user u_i , invoking a metric such as NDCG@K or F1 score.

RUTGERS

Evaluation of Fairness

- Measuring Item-side Fairness: Several measures for evaluating the fairness of a ranked list have been explored in the information retrieval literature [88].
- **Normalized discounted difference (rND)**

rND computes the difference in the proportion of members of the protected group (S+) at top-i and in the over-all population.

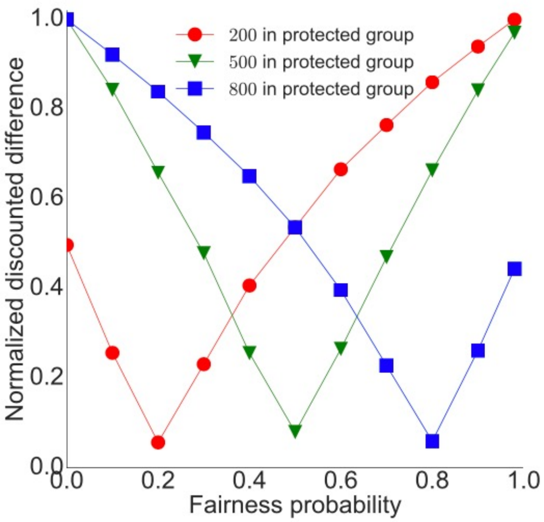
$$\mathrm{rND}(\tau) = \frac{1}{Z} \sum_{i=10,20,\ldots}^{N} \frac{1}{log_2 i} \left| \frac{|S^+_{1\ldots i}|}{i} - \frac{|S^+|}{N} \right|$$

Normalizer Z is computed as the highest possible value of rND for the given number of items N and protected group size |S+|.

Yang et al. "Measuring Fairness in Ranked Outputs". 2016

Evaluation of Fairness

- Measuring Item-side Fairness: Several measures for evaluating the fairness of a ranked list have been explored in the information retrieval literature [88].
- **Normalized discounted difference (rND)**



The figure plots the behavior of rND on synthetic datasets of 1000 items, with 200, 500 and 800 items in S+, as a function of fairness probability.

Yang et al. “Measuring Fairness in Ranked Outputs”. 2016

Evaluation of Fairness

- Measuring Item-side Fairness: Several measures for evaluating the fairness of a ranked list have been explored in the information retrieval literature [88].
- **Normalized discounted KL-divergence (rKL)**

P is the presented exposure distribution;
Q is the desired distribution

$$P = \left(\frac{|S^+_{1\ldots i}|}{i}, \frac{|S^-_{1\ldots i}|}{i} \right), Q = \left(\frac{|S^+|}{N}, \frac{|S^-|}{N} \right)$$

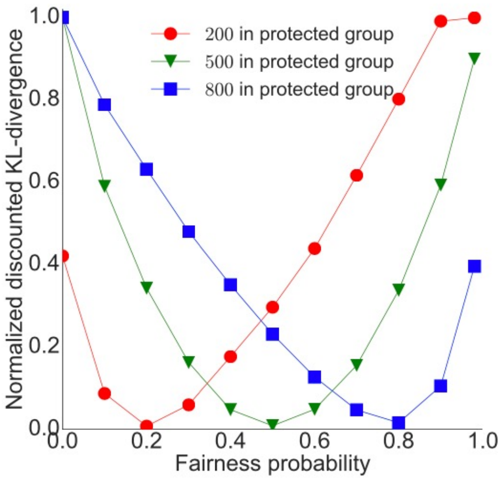
$$\text{rKL}(\tau) = \frac{1}{Z} \sum_{i=10,20,\ldots}^{N} \frac{D_{KL}(P||Q)}{log_2 i}$$

It uses KL-divergence to compute the expectation of the difference between protected group membership at top-i vs. in the over-all population

Yang et al. "Measuring Fairness in Ranked Outputs". 2016

Evaluation of Fairness

- Measuring Item-side Fairness: Several measures for evaluating the fairness of a ranked list have been explored in the information retrieval literature [88].
- **Normalized discounted KL-divergence (rKL)**



The figure plots the behavior of rKL on synthetic datasets of 1000 items, with 200, 500 and 800 items in S+, as a function of fairness probability.

Yang et al. “Measuring Fairness in Ranked Outputs”. 2016

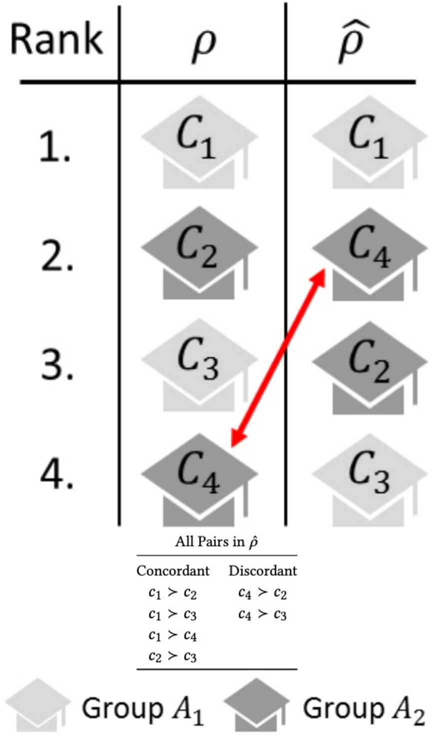
Evaluation of Fairness

- Measuring pairwise fairness by comparing utility and prediction errors [92,93,94]

DEFINITION ***Rank Calibration Error***

$$Rcal_{A_i}(\rho, \hat{\rho}) = \frac{\phi_i^D(X)}{\phi(X) - \phi(A_j)}$$

Where $\phi_i^D(X)$ denotes the number of discordant pairs containing at least one object from the target group A_i.



Kuhlman et al. "Fare: Diagnostics for fair ranking using pairwise error metrics". WWW '19

Evaluation of Fairness

- Measuring Item-side Fairness: Researchers also use Gini Index to do the evaluation at an individual level [84,91].

$$G=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|}{2\sum_{i=1}^{n}\sum_{j=1}^{n}x_j}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|}{2n\sum_{j=1}^{n}x_j}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i-x_j|}{2n^2\bar{x}}$$

Ge et al. “Towards Long-term Fairness in Recommendation”. WSDM’21

Fairness in Recommendation — Dataset

Dataset	#Interactions	Sensitive Features	Reference
ModCloth	99,893	Gender	[50]
RentTheRunway	192,544	Age	[51]
MovieLens	1,000,000	Age; Gender; Occupation	[52]
Insurance	5,382	Gender; Marital status; Occupation	[53]
Post	71,800	Gender	[54]
Coat	11,600	Age; Gender	[55]
Sushi	50,000	Age; Gender	[56]

Challenges and Opportunities

- No Consensus on Definition
- Transparent Fairness
- Fairness-Utility Relationship
- Better Evaluation
- …

Challenges and Opportunities

- No Consensus on Definition
- Transparent Fairness
- Fairness-Utility Relationship
- Better Evaluation
- …

Challenges and Opportunities

No Consensus on Definition

Transparent Fairness

Fairness-Utility Relationship

Better Evaluation

…

Challenges and Opportunities

- No Consensus on Definition
- Transparent Fairness
- Fairness-Utility Relationship
- Better Evaluation
- …

Summary

Introduction and Background:

- Social impact of recommender system and fairness.
- Motivation of fairness.
- Relationship with AI Ethics & Beyond Ethics.

Fairness in Machine Learning:

- Fairness in Classification
- Fairness in Ranking

Fairness in Recommendation:

- Introduction
- Taxonomy
- Dataset and Evaluation
- Challenge and Opportunity

RUTGERS
UNIVERSITY | NEW BRUNSWICK

Questions?

Yunqi Li, Yingqiang Ge, Yongfeng Zhang
Rutgers University
yunqi.li@rutgers.edu
yingqiang.ge@rutgers.edu
yongfeng.zhang@rutgers.edu

Tutorial website: https://fairness-tutorial.github.io/

Reference

- [1] Y. Zhang, Q. Zhao, Y. Zhang, D. Friedman, M. Zhang, Y. Liu, S. Ma. “Economic Recommendation with Surplus Maximization.” WWW 2016.
- [2] Malgieri, Gianclaudio. “The concept of fairness in the GDPR: a linguistic and contextual interpretation.” FAT 2020.
- [3] Faisal Kamiran and Toon Calders. “Classification with No Discrimination by Preferential Sampling.” In BENELEARN, 2010.
- [4] Binh Thanh Luong, Salvatore Ruggieri, and Franco Turini. “kNN as an Implementation of Situation Testing for Discrimination Discovery and Prevention.” In KDD, 2011.
- [5] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. “Certifying and Removing Disparate Impact.” In KDD, 2015.
- [6] Flavio Calmon, Dennis Wei, Bhanukiran Vinzamuri, Karthikeyan Natesan Ramamurthy, and Kush R Varshney. “Optimized Pre-Processing for Discrimination Prevention.” In NIPS, 2017.
- [7] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. “Fairness-aware Classifier with Prejudice Remover Regularizer.” In PADM, 2011.
- [8] Gabriel Goh, Andrew Cotter, Maya Gupta, and Michael Friedlander. “Satisfying Real-world Goals with Dataset Constraints.” In NIPS, 2016.
- [9] Blake Woodworth, Suriya Gunasekar, Mesrob I. Ohannessian, and Nathan Srebro. “Learning Non-Discriminatory Predictors.” In COLT, 2017.

Reference

- [10] Novi Quadrianto and Viktoriia Sharmanska. “Recycling Privileged Learning and Distribution Matching for Fairness.” In NIPS, 2017
- [11] Moritz Hardt, Eric Price, and Nathan Srebro. “Equality of Opportunity in Supervised Learning.” In NIPS, 2016.
- [12] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. “Algorithmic Decision Making and the Cost of Fairness.” In KDD, 2017.
- [13] Aditya Krishna Menon and Robert C Williamson. “The Cost of Fairness in Binary Classification.” FAT 2018.
- [14] A. Asuncion and D.J. Newman. 2007. UCI Machine Learning Repository. (2007). http://www.ics.uci.edu/$\sim$mlearn/
- [15] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. (2017). http://archive.ics.uci.edu/ml
- [16] Joy Buolamwini and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research), Sorelle A. Friedler and Christo Wilson (Eds.), Vol. 81. PMLR, New York, NY, USA, 77–91. http://proceedings.mlr.press/v81/buolamwini18a.html
- [17] Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2018. “Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods.” (2018). arXiv:cs.CL/1804.06876
- [18] M Redmond. 2011. Communities and crime unnormalized data set. UCI Machine Learning Repository. In website: http://www. ics. uci. edu/mlearn/MLRepository. html (2011).

Reference

- [19] J Larson, S Mattu, L Kirchner, and J Angwin. 2016. Compas analysis. GitHub, available at: https://github.com/propublica/compas-analysis[Google Scholar] (2016).
- [20] Manel Capdevila, Marta Ferrer, and Eulália Luque. 2005. La reincidencia en el delito en la justicia de menores. Centro de estudios jurídicos y formación especializada, Generalitat de Catalunya. Documento no publicado (2005).
- [21] Michele Merler, Nalini Ratha, Rogerio S Feris, and John R Smith. 2019. Diversity in Faces. arXiv preprint arXiv:1901.10436 (2019).
- [22] Harald Steck. 2018. “Calibrated recommendations.” In RecSys. 154–1
- [23] Singh, Ashudeep, and Thorsten Joachims. “Policy learning for fairness in ranking.” *arXiv preprint arXiv:1902.04056* (2019)”
- [24] Yadav, Himank, Zhengxiao Du, and Thorsten Joachims. “Fair learning-to-rank from implicit feedback.” *arXiv preprint arXiv:1911.08054* (2019)”
- [25] Narasimhan, Harikrishna, et al. “Pairwise fairness for ranking and regression.” *AAAI* 2020.
- [26] Sara Hajian, Francesco Bonchi, and Carlos Castillo. “Algorithmic bias: From discrimination discovery to fairness-aware data mining.” SIGKDD 2016.
- [27] Biega, A. J. et al. “Equity of Attention: Amortizing Individual Fairness in Rankings” SIGIR’18.
- [28] Zehlike, Meike, et al. “Fa*ir: A fair top-k ranking algorithm.” *CIKM* 2017.

RUTGERS

Reference

- [29] AirBnB. https://insideairbnb.com
- [30] Correctional Offender Management Profiling for Alternative Sanctions. https://github.com/propublica/compas-analysis
- [31] CSRankings: Computer Science Rankings. https://csrankings.org
- [32] Engineering students. https://github.com/MilkaLichtblau/DELTR-Experiments/tree/master/data/EngineeringStudents
- [33] SAT. https://www.qsleap.com/sat/resources/sat-2014-percentiles
- [34] German Credit. https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/
- [35] Forbes Richest Americans. https://github.com/DataResponsibly/Datasets
- [36] XING. https://github.com/MilkaLichtblau/xing_dataset
- [37] Fu, Zuohui, et al. "Fairness-aware explainable recommendation over knowledge graphs." *SIGIR*. 2020.
- [38] Himan Abdollahpouri and Robin Burke. 2019. "Multi-stakeholder recommendation and its connection to multi-sided fairness." arXiv preprint arXiv:1907.13158(2019)
- [39] Robin Burke. 2017. "Multisided fairness for recommendation." arXiv preprintarXiv:1707.00093(2017)
- [40] Gourab K Patro, et al. "FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms." WWW 2020.
- [41] Jurek Leonhardt, Avishek Anand, and Megha Khosla. 2018. "User Fairness in Recommender Systems." WWW 2018.

RUTGERS

Reference

- [42] Himan Abdollahpouri, et al. "The unfairness of popularity bias in recommendation." arXiv preprintarXiv:1907.13286(2019)
- [43] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. "Controlling popularity bias in learning-to-rank recommendation." RecSys. 42–46
- [44] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. "Managing popularity bias in recommender systems with personalized re-ranking." arXivpreprint arXiv:1901.07555(2019).
- [45] Gediminas Adomavicius and YoungOk Kwon. 2011. "Improving aggregate recommendation diversity using ranking-based techniques." TKDE(2011)
- [46] Aria Khademi, Sanghack Lee, David Foley, and Vasant Honavar. "Fairness in algorithmic decision making: An excursion through the lens of causality." WWW 2019. 2907–2914
- [47] Matt Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. "Counterfactual fairness". In NeurIPS. 2017
- [48] Junzhe Zhang and Elias Bareinboim. 2018. "Equality of opportunity in classification: A causal approach." In NeurIPS. 3675–3685.
- [49] Junzhe Zhang and Elias Bareinboim. 2018. "Fairness in decision-making—the causal explanation formula." InAAAI, Vol. 32
- [50] Mengting Wan, Jianmo Ni, Rishabh Misra, Julian McAuley "Addressing Marketing Bias in Product Recommendations". WSDM, 2020

Reference

- [51] Rishabh Misra, Mengting Wan, Julian McAuley, "Decomposing fit semantics for product size recommendation in metric spaces" RecSys, 2018
- [52] https://grouplens.org/datasets/movielens/
- [53] https://www.kaggle.com/mrmorj/insurance-recommendation
- [54] https://www.kaggle.com/vatsalparsaniya/post-pecommendation
- [55] Schnabel, Tobias, et al. "Recommendations as treatments: Debiasing learning and evaluation." *international conference on machine learning*. PMLR, 2016.
- [56] Toshihiro Kamishima. "Nantonac collaborative filtering: recommendation based on order responses." SIGKDD 2003.
- [57] Toshihiro Kamishima, et al. "Fairness-aware classifier with prejudice remover regularizer." In Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2012
- [58] Aditya Krishna Menon, et al. "The cost of fairness in binary classification." FAT 2018
- [59] Emmanouil Krasanakis, et al. "Adaptive Sensitive Reweighting to Mitigate Bias in Fairness-aware Classification". WWW 2018.
- [60] Berk Ustun, et al. "Fairness without Harm: Decoupled Classifiers with Preference Guarantees." ICML 2019
- [61] Muhammad Bilal Zafar et al. "Fairness constraints: Mechanisms for fair classification." arXiv:1507.05259 (2015).
- [62] Yongkai Wu, et al. "Fairness-aware Classification: Criterion, Convexity, and Bounds." arXiv:cs.LG/1809.04737 (2018).

RUTGERS

Reference

- [63] Lingxiao Huang and Nisheeth Vishnoi. "Stable and Fair Classification. " ICML 2019
- [64] Toon Calders and Sicco Verwer. "Three naive Bayes approaches for discrimination-free classification." Data Mining and Knowledge Discovery 21, 2 (2010), 277–292.
- [65] Biega, Asia J.,et al. "Equity of attention: Amortizing individual fairness in rankings." *SIGIR* 2018.
- [66] Zehlike, Meike, et al. "Fa*ir: A fair top-k ranking algorithm." *CIKM* 2017.
- [67] Singh, Ashudeep, and Thorsten Joachims. "Policy learning for fairness in ranking." *arXiv preprint arXiv:1902.04056* (2019).
- [68] Yadav, Himank, et al. "Fair learning-to-rank from implicit feedback." *arXiv preprint arXiv:1911.08054* (2019).
- [69] Yang, Ke, and Julia Stoyanovich. "Measuring fairness in ranked outputs." *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. 2017.
- [70] Narasimhan, Harikrishna, et al. "Pairwise fairness for ranking and regression." *AAAI* 2020.
- [71] Celis, L. Elisa, et al. "Ranking with fairness constraints." *arXiv preprint arXiv:1704.06840* (2017).
- [72] Zehlike, Meike, and Carlos Castillo. "Reducing disparate exposure in ranking: A learning to rank approach." *Proceedings of The Web Conference 2020.*
- [73] Vogel, Robin, et al. "Learning Fair Scoring Functions: Fairness Definitions, Algorithms and Generalization Bounds for Bipartite Ranking." *arXiv preprint arXiv:2002.08159* (2020).

Reference

- [74] “Zafar, Muhammad Bilal, et al. "Fairness Constraints: A Flexible Approach for Fair Classification." *J. Mach. Learn. Res.* 20.75 (2019): 1-42.
- [75] Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *SIGKDD*’2018.
- [76] Li.Y et al. “User-oriented Fairness in Recommendation” WWW’21.
- [77] Xiao, Lin, et al. "Fairness-aware group recommendation with pareto-efficiency." Recsys’17
- [78] Beutel, Alex, et al. “Fairness in Recommendation Ranking through Pairwise Comparisons” SIGKDD’19
- [79] Yao, Sirui, and Bert Huang. “Beyond Parity: Fairness Objectives for Collaborative Filtering” NIPS’17
- [80] Li. Y et al. “Towards Personalized Fairness based on Causal Notion” SIGIR’21
- [81] Sühr,Tom,et al.“Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform." ACM KDD 2019.
- [82] Patro, Gourab, et al.“FairRec: Two-sided Fairness for Personalized Recommendations in Two-Sided Platforms." WWW 2020.
- [83] Yao Wu, Jian Cao, Guandong Xu, and Yudong Tan. 2021. “TFROM: A Two-sided Fairness-Aware Recommendation Model for Both Customers and Providers.” arXiv:2104.09024 [cs.AI]
- [84] Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. 2021. “Multi-FR:A Multi-Objective Optimization Method for Achieving Two-sided Fairness in E-commerce Recommendation. ” arXiv:2105.02951 [cs.IR]

RUTGERS

Reference

- [85] Nyi Nyi Htun, Elisa Lecluse, and Katrien Verbert. 2021. “Perception of Fairness in Group Music Recommender Systems.” In26th International Conference on Intelligent User Interfaces (College Station, TX, USA)(IUI ’21). Association for Computing Machinery, New York, NY, USA, 302–306.
- [86] Sriram Vasudevan and Krishnaram Kenthapadi. 2020. “The LinkedIn FairnessToolkit” (LiFT)
- [87] Geyik, S. C., Ambler, S., & Kenthapadi, K. (2019). “Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search.” In Proceedings of the 25th ACM SIGKDD International Con-805ference on Knowledge Discovery & Data MiningKDD ’19 (pp. 2221–2231).New York, NY, USA: ACM.
- [88] K. Yang and J. Stoyanovich. “Measuring fairness in ranked outputs.” In SSDBM, 2017.
- [89] Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., Xu, S., Geng, S., Shah, C., Zhang, Y. and De Melo, G., 2020, July.” Fairness-aware explainable recommendation over knowledge graphs.” In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 69-78).
- [90] Wang, G., Zhang, Y., Fang, Z., Wang, S., Zhang, F. and Zhang, D., 2020. “FairCharge: A data-driven fairness-aware charging recommendation system for large-scale electric taxi fleets.” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(1), pp.1-25.
- [91] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, and Yongfeng Zhang. 2021. “Towards Long-term Fairness in Recommendation.” In Proceedings of the 14th ACM Conference on Web Search and Data Mining (WSDM).

Reference

- [92] Kuhlman, C., Van Valkenburg, M., & Rundensteiner, E. (2019). ”Fare: Diagnostics for fair ranking using pairwise error metrics.” In The World Wide Web Conference WWW ’19 (pp. 2936–2942).New York, NY, USA: ACM.
- [93] Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E. H., & Goodrow, C. (2019). “Fairness in recommendation ranking through pairwise comparisons.” In Proceedings of the 25th ACMSIGKDD International Conference on Knowledge Discovery & Data Mining KDD ’19 (p. 2212–2220). New York, NY, USA: Association for Computing Machinery.
- [94] Yao, S., & Huang, B. (2017b). “New fairness metrics for recommendation that embrace differences.” arXiv preprint arXiv:1706.09838.
- [95] Castelnovo, Alessandro, et al. "The zoo of Fairness metrics in Machine Learning." *arXiv preprint arXiv:2106.00467* (2021).