推荐系统中的公平性问题

刘卫文

香港中文大学

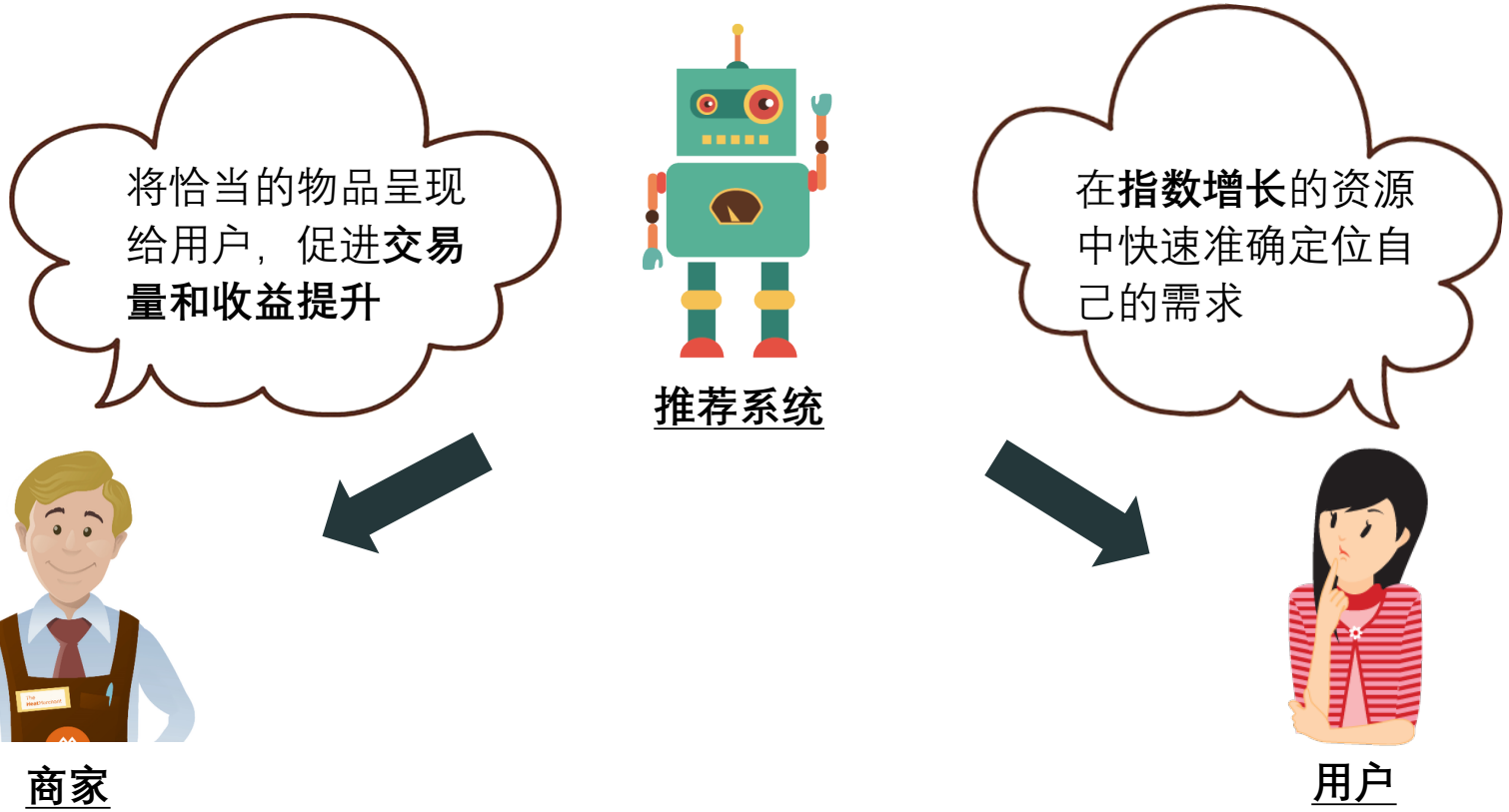
wwliu@cse.cuhk.edu.hk

目录

- 推荐系统中的公平性问题
  - 公平性问题的来源
  - 公平性问题的重要性
  - 公平性问题的分类
  - 公平性问题的形式
  - 准确率和公平性的权衡
- 解决方法
  - 商品消费者的公平性标准[NIPS'17]
  - 带公平性约束的优化问题[KDD'18]
  - 校准推荐[RecSys'18]
  - 公平的贷款推荐算法(PFAR)[RecSys'18]
  - 合作项目:使用强化学习进行准确性和公平性的权衡
- 未来的研究方向

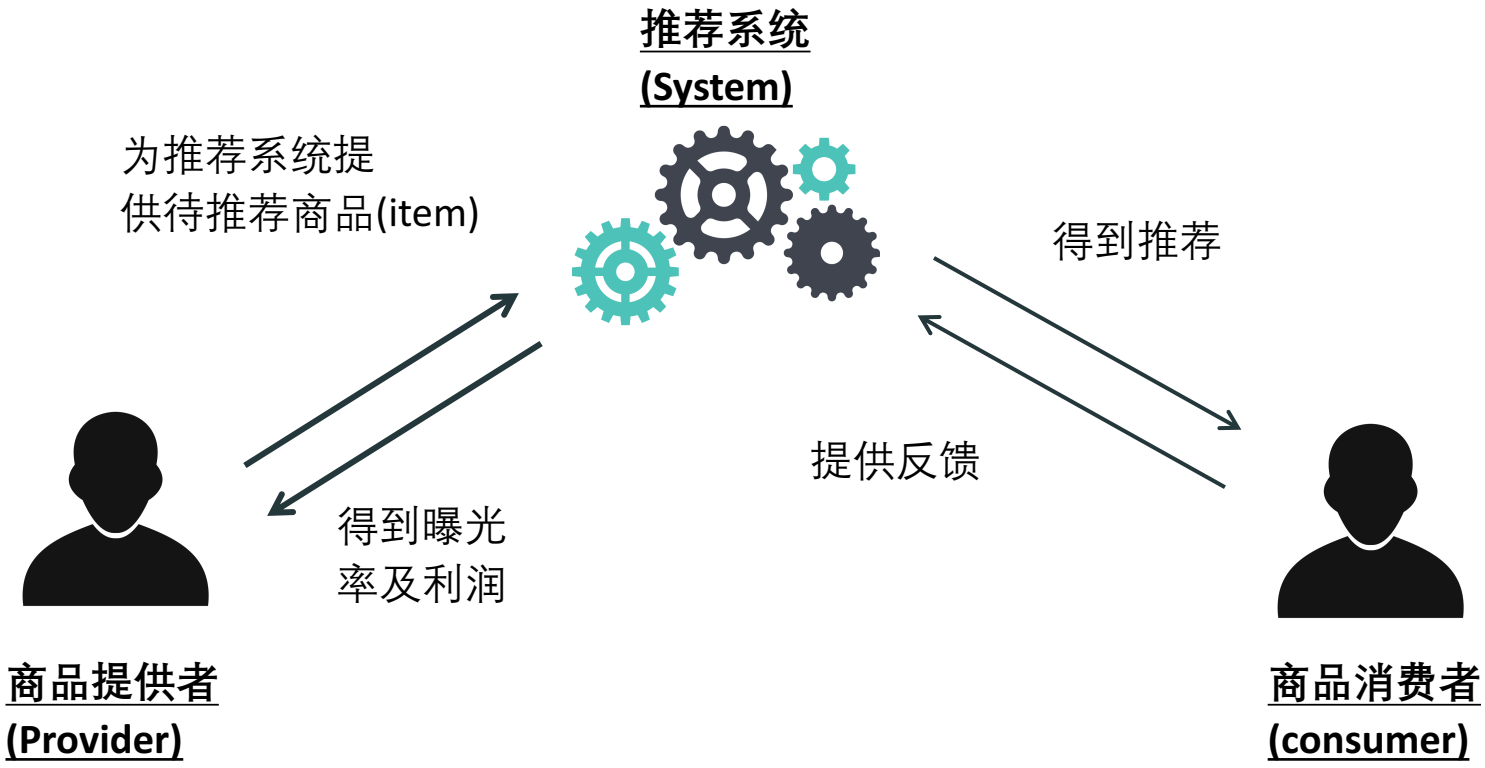
推荐系统

- 推荐系统是一种信息过滤系统，能根据用户的档案或者历史行为记录，学习出用户的兴趣爱好，预测出用户对给定物品的评分或者偏好。



多主体推荐系统(Multisided Recommender Systems)

- 实际应用场景中，推荐系统往往需要考虑多方的利益，有多个优化主体



实际生活中的公平性问题

BBC Sign in | News | Sport | Reel | Worklife | Travel | Future

NEWS

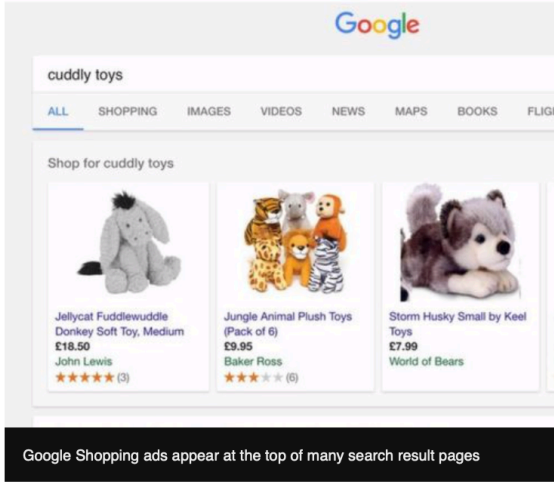
Home | Video | World | Asia | UK | Business | Tech | Science | Stories | Entertainmer

Technology

Google hit with record EU fine over Shopping service

By Leo Kelion
Technology desk editor

27 June 2017

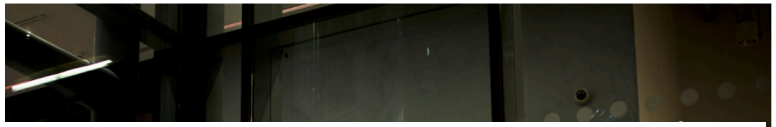


Google Shopping ads appear at the top of many search result pages

Google has been fined 2.42bn euros ($2.7bn; £2.1bn) by the Eur Commission after it ruled the company had abused its power by its own shopping comparison service at the top of search result

The New York Times

Google Fined $1.7 Billion by E.U. for Unfair Advertising Rules



Google Image search for CEO has Barbie as first female result

By Amelia Butterly
Newsbeat reporter

TECH | 16 Apr 2015

Search the term "CEO" in Google Images and the first picture of woman you get is a picture of Barbie in a suit.

Most Popular

实际生活中的公平性问题

Unequal Representation and Gender Stereotypes in Image Search Results for Occupations

Matthew Kay
Computer Science & Engineering | dub, University of Washington
mjskay@uw.edu

Cynthia Matuszek
Computer Science & Electrical Engineering, University of Maryland Baltimore County
cmat@umbc.edu

Sean A. Munson
Human-Centered Design & Engineering | dub, University of Washington
smunson@uw.edu

ABSTRACT
Information environments have the power to affect people's perceptions and behaviors. In this paper, we present the results of studies in which we characterize the gender bias present in image search results for a variety of occupations. We experimentally evaluate the effects of bias in image search results on the images people choose to represent those careers and on people's perceptions of the prevalence of men and women in each occupation. We find evidence for both stereotype exaggeration and systematic underrepresentation of women in search results. We also find that people rate search results higher when they are consistent with stereotypes for a career, and shifting the representation of gender in image search results can shift people's perceptions about real-world distributions. We also discuss tensions between desires for high-quality results and broader societal goals for equality of representation in this space.

Author Keywords
Representation; bias; stereotypes; gender; inequality; image search

tional choices, opportunities, and compensation [20,26]. Stereotypes of many careers as gender-segregated serve to reinforce gender sorting into different careers and unequal compensation for men and women in the same career. Cultivation theory, traditionally studied in the context of television, contends that both the prevalence and characteristics of media portrayals can develop, reinforce, or challenge viewers' stereotypes [29].

Inequality in the representation of women and minorities, and the role of online information sources in portraying and perpetuating it, have not gone unnoticed in the technology community. This past spring, Getty Images and LeanIn.org announced an initiative to increase the diversity of working women portrayed in the stock images and to improve how they are depicted [27]. A recent study identified discrimination in online advertising delivery: when searching for names, search results for black-identifying first names were accompanied by more ads for public records searches than those for white-identifying first names, and those results were more likely to suggest searches for arrest records [34].

工作推荐中的公平性问题

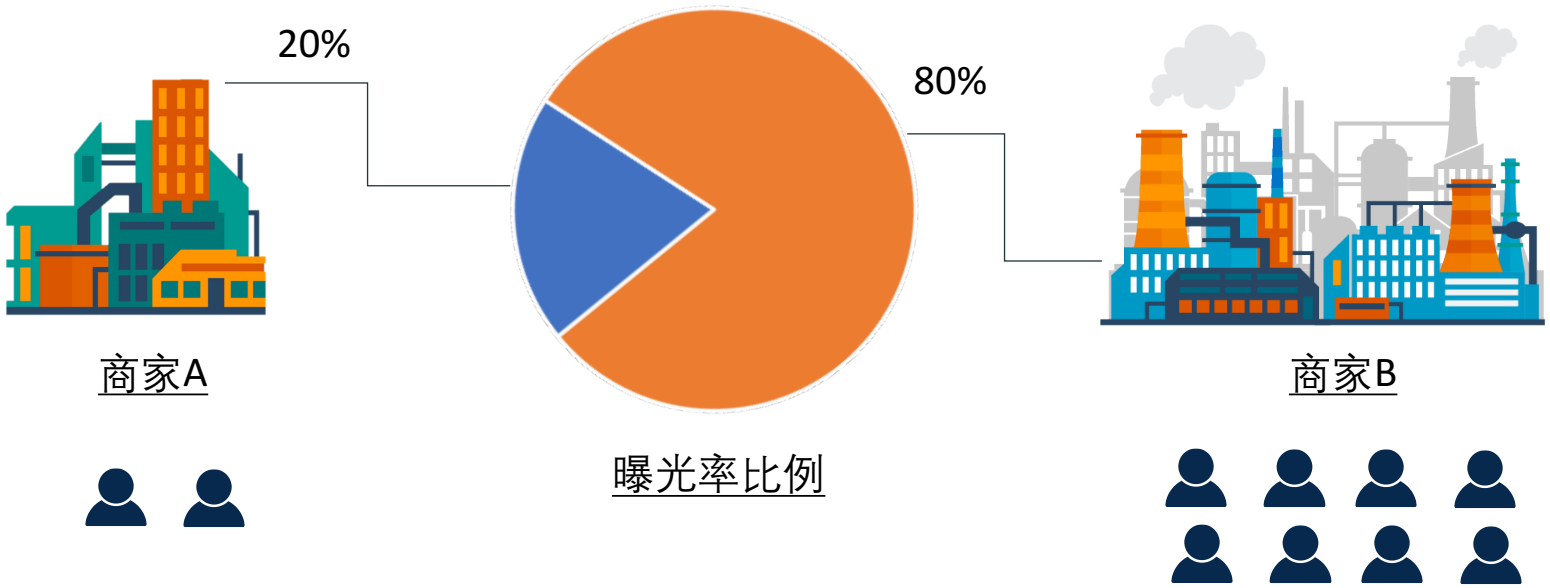
- 能力相似的申请者，因为性别/年龄/种族等，得到的推荐结果（如薪资级别，推荐质量）有差别。

申请人	推荐结果
	2k
	3k
	1.6k
	1.3k
	1k
	2k

申请人	推荐结果
	8k
	7k
	6k
	5.3k
	7.1k
	9k

电商平台中的公平性问题

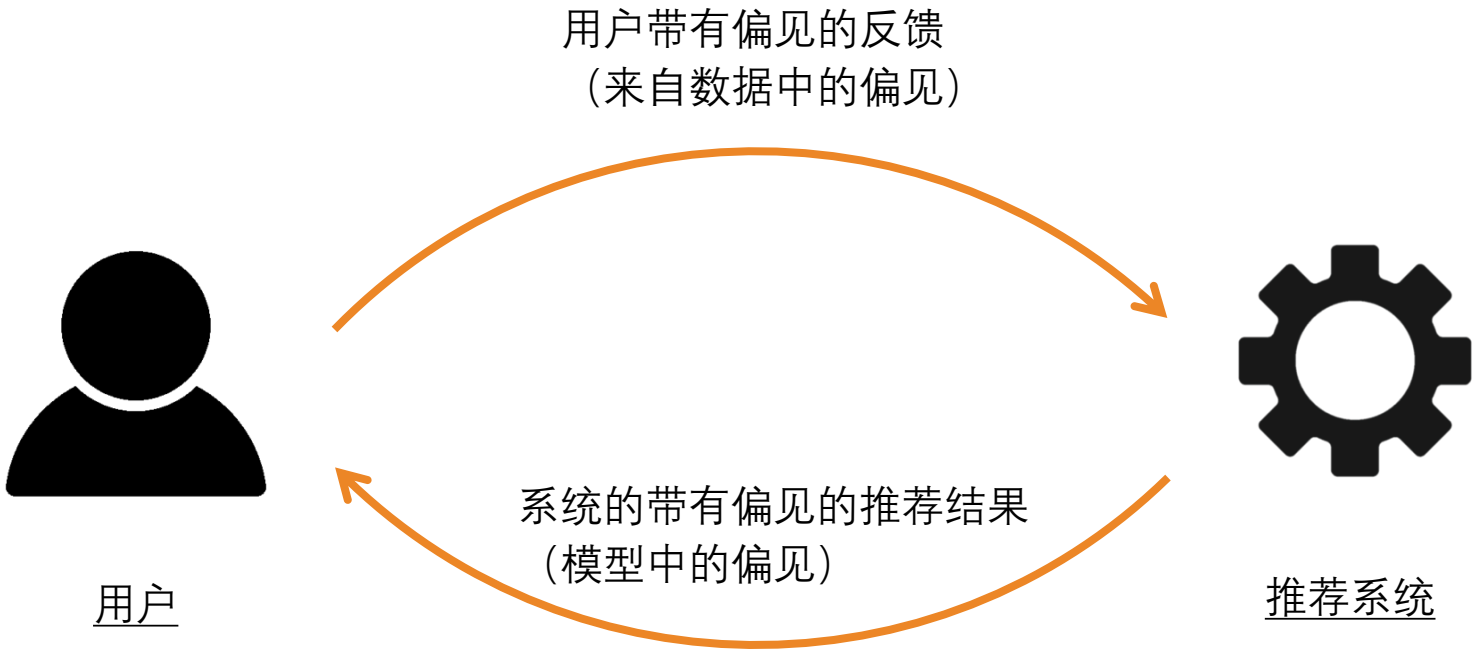
- 商品质量相同的店家，因为商家规模/成立时间长短/地区等，得到的曝光率（被展示给用户的次数）有差别。



公平性问题的来源

- 数据
  - 带有偏见的用户行为（评论，浏览点击历史等）
  - 数据收集时，不同群体被收集的数据量不同。算法往往在拥有大数据量的群体上取得更好的学习效果（“多数人的暴政”/popularity bias）
- 算法
  - 算法使用敏感属性进行学习
  - 错把少数群体当作噪声处理
  - 使用未考虑公平性问题的评价指标

公平性问题的来源



公平性问题的重要性

- **法律上**
  - 出台反歧视法顺应社会发展趋势。
    - 我国已出台《反就业歧视法》
    - 欧盟《一般数据保护条例》(GDPR)提出了算法公平性原则
    - 美国《反歧视法》提出禁止就业，住房，教育和其他社会生活领域的歧视。
- **经济上**
  - 解决公平性问题有利于系统的长期健康发展。反之，受到差别待遇的群体倾向于放弃该系统，影响长期效益。
- **公共关系上**
  - 解决公平性问题有利于建立用户对于系统的信心，获得良好声誉。

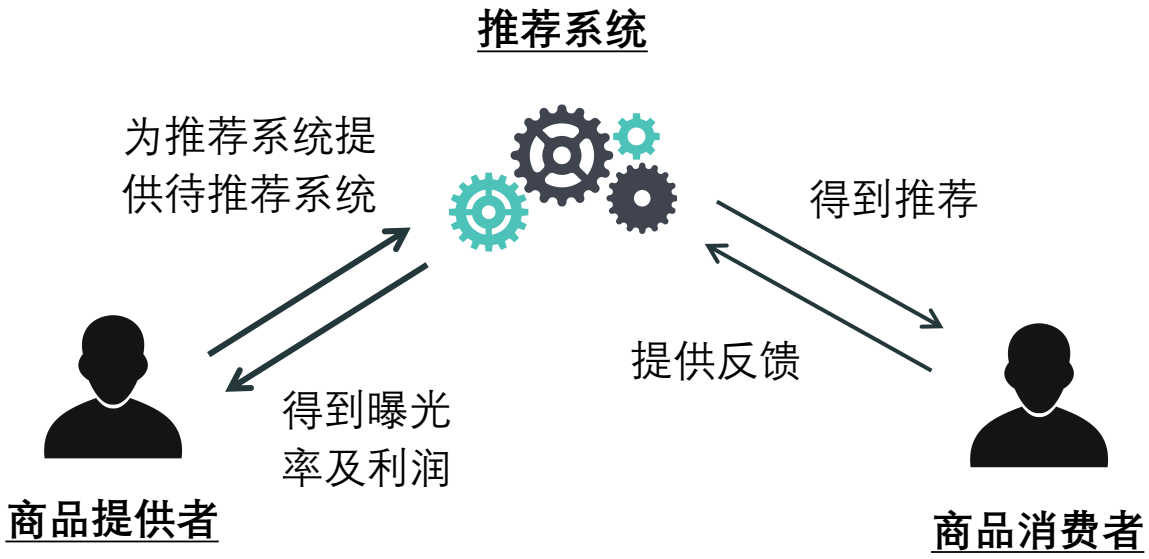
亚当斯的公平理论

- **Equity theory** focuses on determining whether the distribution of resources is fair to both relational partners. Equity is measured by comparing the ratio of contributions (or costs) and benefits (or rewards) for each person.
- 公平理论侧重于确定资源分配对两个关系伙伴是否公平。公平是通过比较每个人的贡献（或成本）与收益（或奖励）的比率来衡量的。

------- J. Stacy Adams, 1960

公平性问题的分类：按主体

- 一个典型的推荐生态系统包括了三种身份：商品提供者，商品消费者，和推荐系统
  - **商品消费者侧的公平性问题(Consumer-side Fairness)**
    - 不同群体得到的推荐效果是否公平
  - **商品提供者侧的公平性问题(Provider-side Fairness)**
    - 不同群体曝光率分配是否均衡



公平性问题的分类：按粒度

- **个人公平性**
  - 能力相似的个人应该得到相似的对待
- **群体公平性**
  - 不同群体应该得到相似的对待，如黑人群体，贫困群体

公平性标准

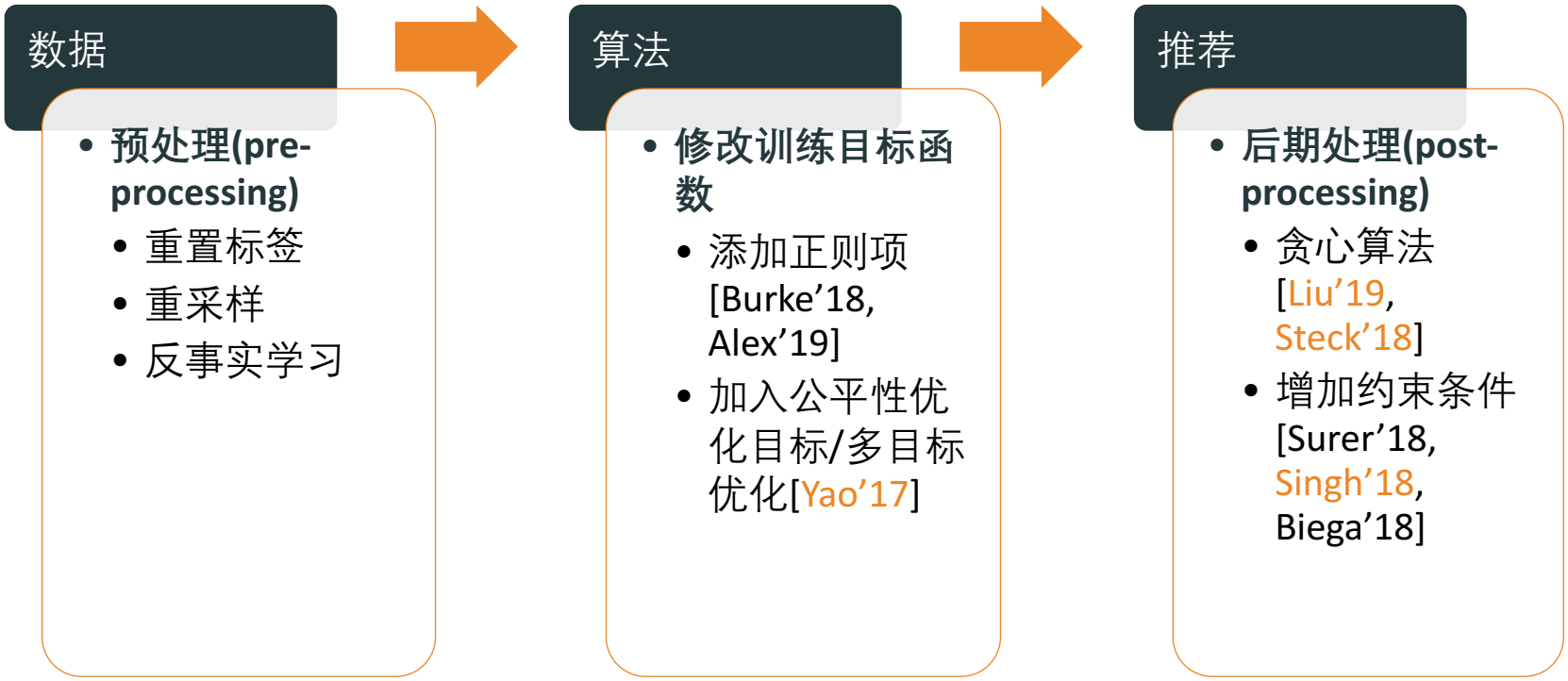
- **统计平等(Demographic Parity/Statistical Parity)**
  - 指受保护群体与其他群体享有同等机会(equality of opportunity)。
- **差别对待(Disparate Treatment)**
  - 差别对待指的是对待受保护群体的方式与另一群体不同（存在因群体受保护特征而导致的歧视）。
- **差别效果(Disparate Impact)**
  - 在雇佣，住房和其他领域中，差别效果指的是对一群受保护群体的不利影响要大于对另一群体的不利影响，即使雇主或者房东所采用的规则在形式上时中立的。

准确率和公平性的权衡

- 推荐准确率
  - 准确预测用户的个性化兴趣
  - 传统推荐系统通常喜欢推荐大众流行的的热门商品，忽略小众商品，使得推荐分布失衡。
- 公平性
  - 较为均匀的将资源进行分配。
  - 强调群体性，弱化了个性化兴趣

算法需要考虑推荐准确率和公平性之间的权衡

解决方法



[Burke'18] Burke, Robin, et al. "Balanced neighborhoods for multi-sided fairness in recommendation." *FAT*, 2018.
[Alex'19] Alex Beutel (Google). "Fairness in Recommendation Ranking through Pairwise Comparisons." *KDD*, 2019.
[Yao'17] Yao, Sirui, and Bert Huang. "Beyond parity: Fairness objectives for collaborative filtering." *NIPS*, 2017.
[Liu'19] Liu, Weiwen, et al. "Personalized fairness-aware re-ranking for microlending." *RecSys*, 2019.
[Steck'18] Steck, Harald. "Calibrated recommendations." *RecSys*, 2018.
[Surer'18] Sürer, Özge, et al. "Multistakeholder recommendation with provider constraints." *RecSys*, 2018.
[Singh'18] Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *KDD*, 2018.
[Biega'18] Biega, Asia J., et al. "Equity of attention: Amortizing individual fairness in rankings." *SIGIR*, 2018.

商品消费者的公平性标准[NIPS'17]

- 希望受保护的用户与其他用户拥有相近的推荐准确率
  - **Value Fairness**

$$U_{\text{val}} = \frac{1}{n}\sum_{v=1}^{n} \left|(E_u[y]_v - E_u[r]_v) - (E_{\neg u}[y]_v - E_{\neg u}[r]_v)\right|.$$

  - **Underestimation Fairness**

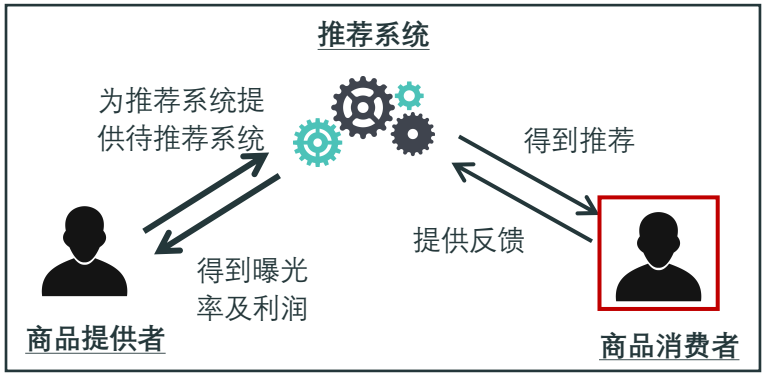
$$U_{\text{under}} = \frac{1}{n}\sum_{v=1}^{n} \left|\max\{0, E_u[r]_v - E_u[y]_v\} - \max\{0, E_{\neg u}[r]_v - E_{\neg u}[y]_v\}\right|.$$

  - **Non-parity Fairness**

$$U_{\text{par}} = \frac{1}{n}\sum_{v=1}^{n} \left|E_u[y]_v - E_{\neg u}[y]_v\right|.$$

- 优化目标: $\min J + U$

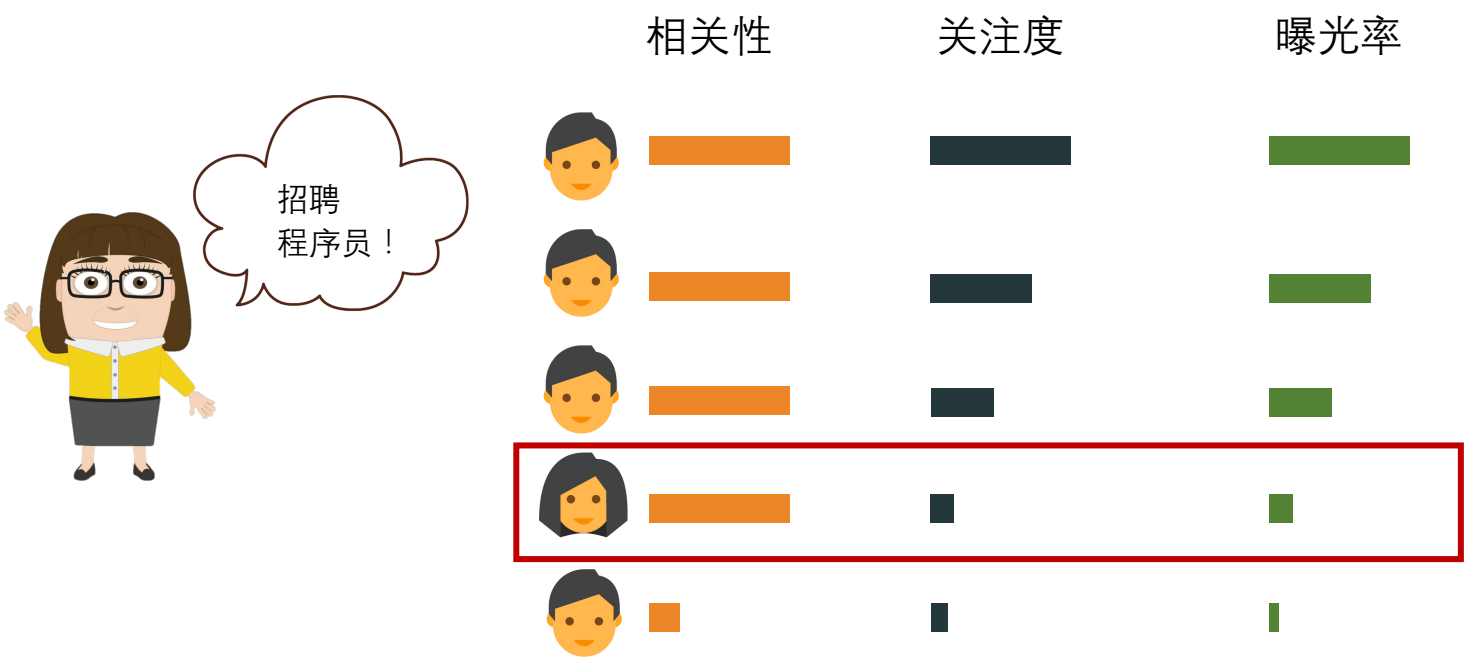
基于传统的推荐系统，忽略了商品提供者的公平性问题。



[Yao'17] Yao, Sirui, and Bert Huang. "Beyond parity: Fairness objectives for collaborative filtering." *NIPS*, 2017.

带公平性约束的优化问题[KDD'18]

- 用户给予列表中商品的关注度由上到下指数递减，与商品相关性无关



[Singh'18] Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *KDD*, 2018.

带公平性约束的优化问题[KDD'18]

- 将公平性问题建模成为一个带公平性约束的优化问题

$$r = \text{argmax}_r U(r|q)$$
$$\text{s.t. } r \text{ is fair}$$

- 常用的效用函数(utility function)

$$U(r|q) = \sum_u P(u|q) \sum_d v\big(\text{rank}(d|r)\big)\lambda\big(\text{rel}(d|u,q)\big) \ .$$

- 公平性约束

- **Demographic Parity:**
- 不同群体的平均曝光率相同

$$\text{Exposure}(G_0|P) = \text{Exposure}(G_1|P).$$

- **Disparate Treatment:**
- 不同群体的平均曝光率与平均效用成相同比例

$$\frac{\text{Exposure}(G_0|P)}{U(G_0|q)} = \frac{\text{Exposure}(G_1|P)}{U(G_1|q)}.$$

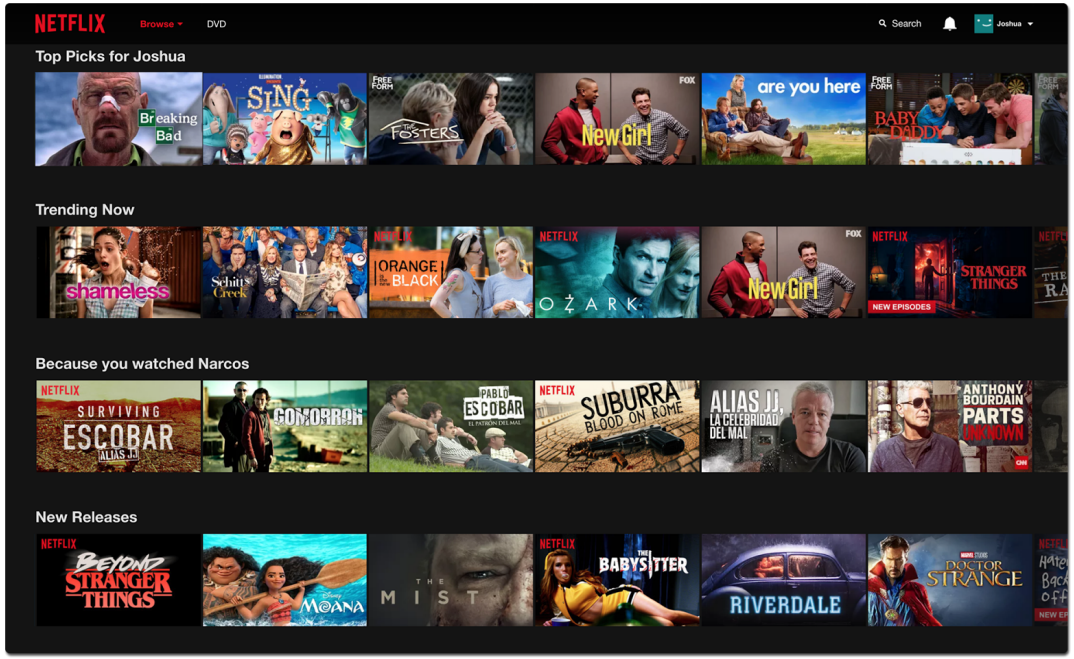
- **Disparate Impact:**
- 不同群体的平均点击率与平均效用成相同比例

$$\frac{CTR(G_0|P)}{U(G_0|q)} = \frac{CTR(G_1|P)}{U(G_1|q)}.$$

[Singh'18] Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *KDD*, 2018.

校准推荐[RecSys'18]

- 公平的推荐结果分布应该与其原始兴趣分布一致⇒通过校准跳出恶性循环
  - 用户历史总共看过70部动作片，30部纪录片，推荐结果也应该有70%是动作片，30%纪录片。而非只注重动作片的推荐，忽略了小众纪录片。



[Steck'18] Steck, Harald. "Calibrated recommendations." *RecSys*, 2018.

校准推荐[RecSys'18]

- 后期处理方法(post-processing)
- 对于推荐结果重新排序
- 每次选取分数最高的i_*进行推荐。其中i是商品，$s(i)$是传统推荐算法预测的打分，p, q分别是已推荐给用户的商品种类分布，和历史种类分布，使用参数λ进行权重调节

$$i_* = \text{argmax}_i \underbrace{(1-\lambda)s(i)}_{\text{准确率}} + \underbrace{\lambda \text{KL}(p||q)}_{\text{校准差异}}$$

[Steck'18] Steck, Harald. "Calibrated recommendations." *RecSys*, 2018.

公平的贷款推荐算法(PFAR)[RecSys'19]

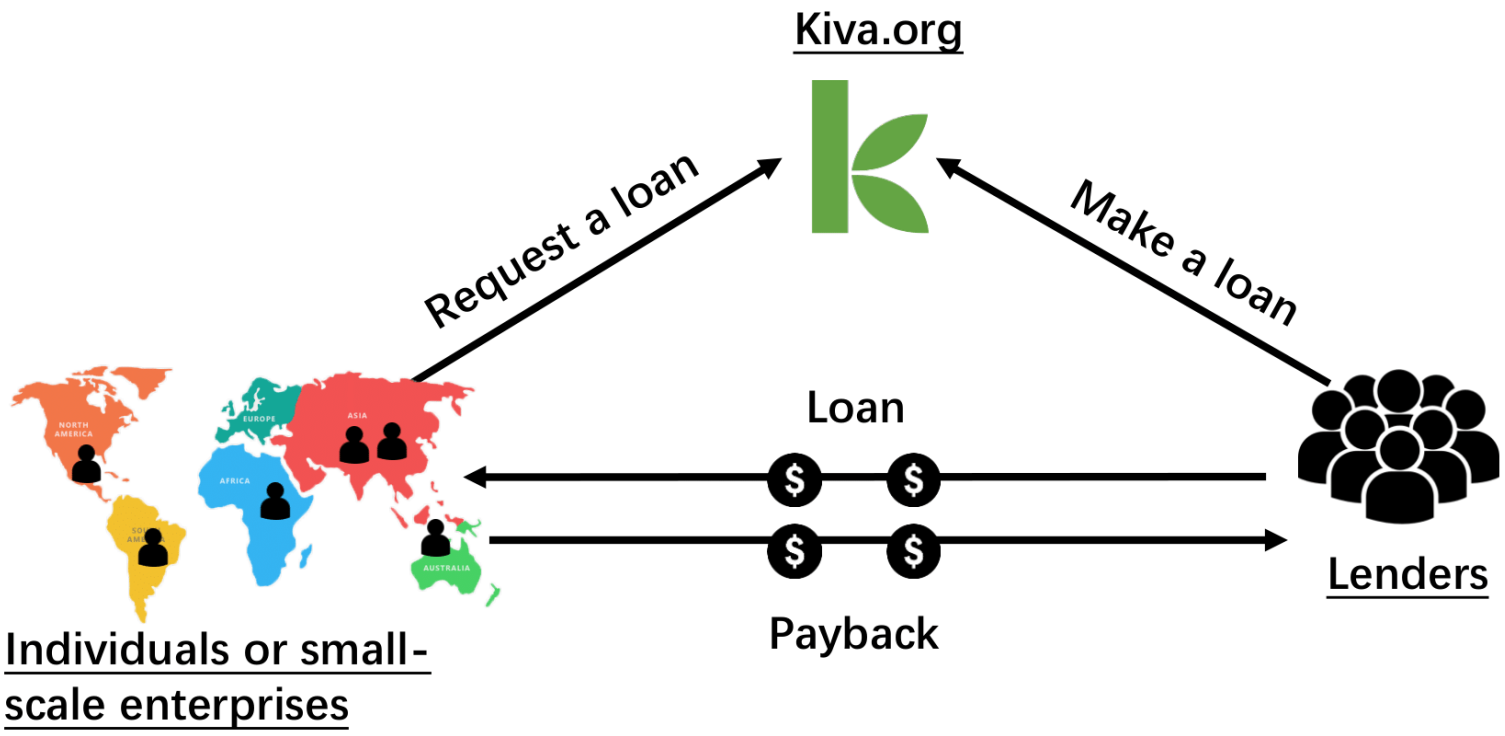


图1. Kiva是一个公益性组织，贷款人向需要贷款的个人或初创公司提供无息贷款。

[Liu'19] Liu, Weiwen, et al. "Personalized fairness-aware re-ranking for microlending." *RecSys*, 2019.

传统推荐算法中的公平性问题

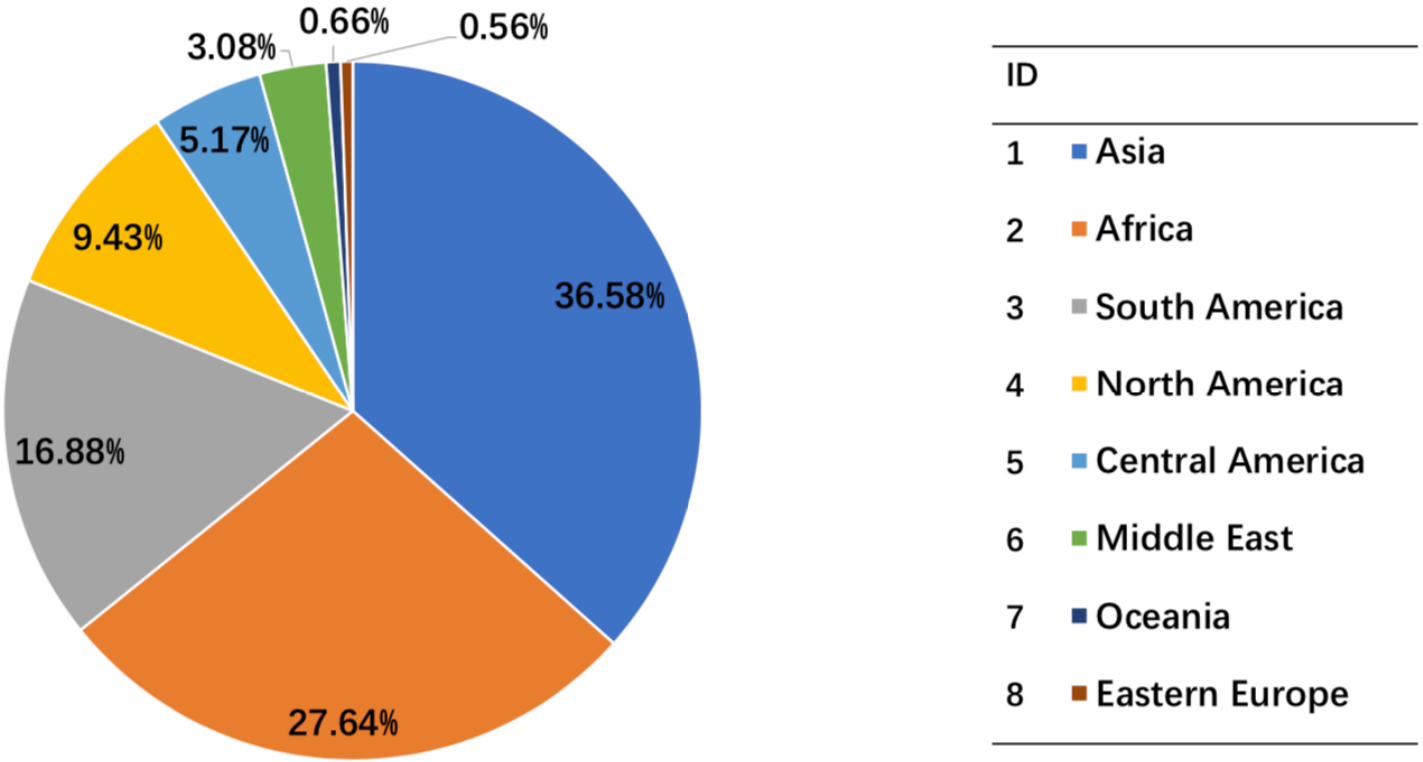


图2. 使用传统算法，来自不同地区的贷款申请得到的推荐次数（曝光率）

PFAR算法

Algorithm 1 (Personalized) Fairness-Aware Re-ranking (FAR/PFAR)

Require: $u, R(u), K, \lambda, \tau_u$

Ensure: $S(u)$

1: $S(u) \leftarrow \emptyset$

2: **while** $|S(u)| < K$ **do**

3: Select the optimal v^* by solving

$$\arg\max_{v \in R(u)} \underbrace{(1-\lambda)P(v|u)}_{\text{personalization}} + \underbrace{\lambda\tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}}_{\text{personalized fairness}}$$

4: $R(u) \leftarrow R(u) \setminus \{v^*\}$

5: $S(u) \leftarrow S(u) \cup \{v^*\}$

6: **end while**

7: **return** $S(u)$

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0.5 = 0.85$
v_3	0.60	1	0	$0.60 + 0.4 \times 0.5 = 0.8$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

ϕ

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0.5 = 0.85$
v_3	0.60	1	0	$0.60 + 0.4 \times 0.5 = 0.8$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

v_1

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0 = 0.65$
v_3	0.60	1	0	$0.60 + 0.4 \times 0 = 0.6$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

v_1

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0 = 0.65$
v_3	0.60	1	0	$0.60 + 0.4 \times 0 = 0.6$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

v_1

v_4

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0 = 0.65$
v_3	0.60	1	0	$0.60 + 0.4 \times 0 = 0.6$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

v_1

v_4

v_2

PFAR算法

$$P(v|u) + \tau_u \sum_c P(\mathcal{V}_c) 1_{\{v \in \mathcal{V}_c\}} \prod_{i \in S(u)} 1_{\{i \notin \mathcal{V}_c\}}$$

- 假设存在两个商品群体$P(\mathcal{V}_1) = P(\mathcal{V}_2) = 0.5, \tau_u = 0.4.$

商品	$P(v\|u)$	$\mathcal{V}_1$	$\mathcal{V}_2$	公平性得分
v_1	0.70	1	0	$0.70 + 0.4 \times 0.5 = 0.9$
v_2	0.65	1	0	$0.65 + 0.4 \times 0 = 0.65$
v_3	0.60	1	0	$0.60 + 0.4 \times 0 = 0.6$
v_4	0.55	0	1	$0.55 + 0.4 \times 0.5 = 0.75$

输出重排列表S:

v_1

v_4

v_2

v_3

PFAR实验结果

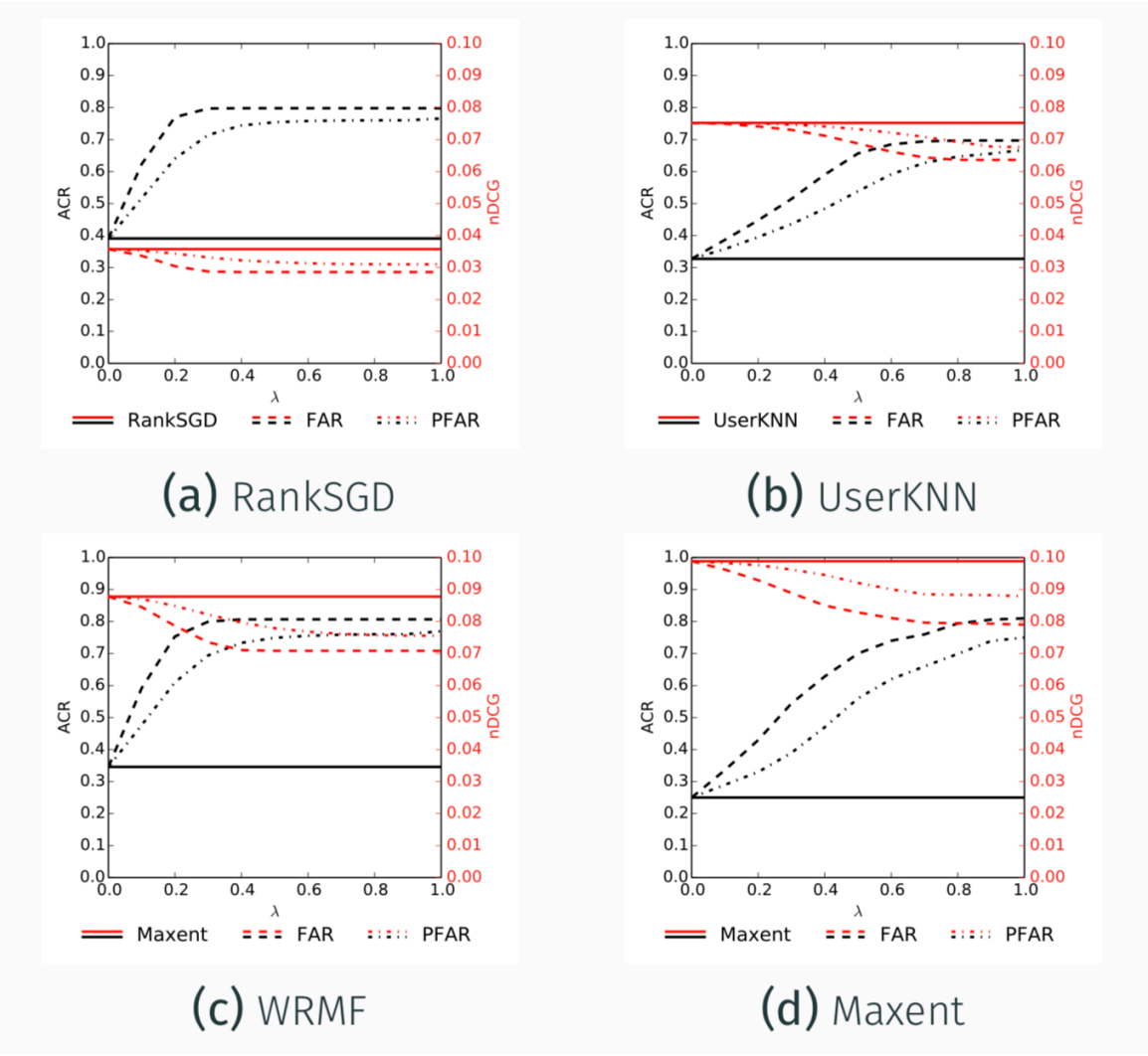


图3. 将FAR和PFAR应用到四种传统推荐算法，在Kiva数据集上的效果

PFAR实验结果

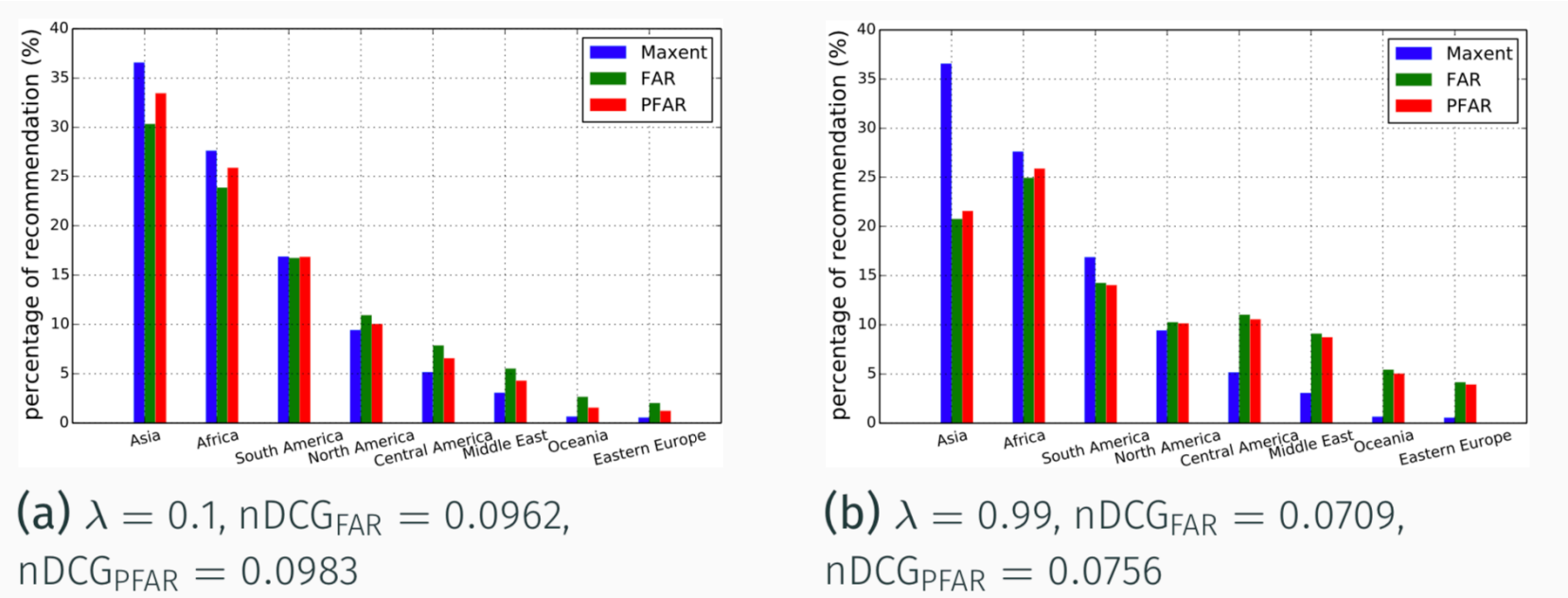
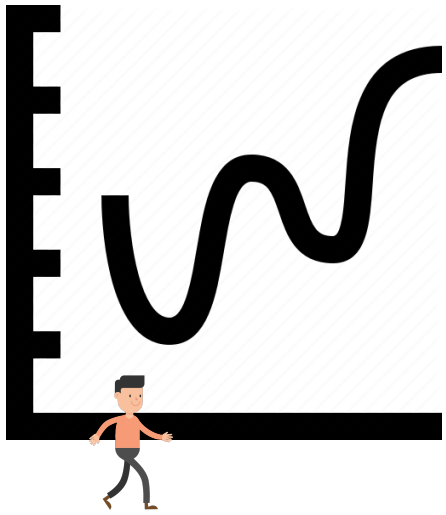


图4. 将FAR和PFAR应用到贷款推荐上，各地区贷款申请得到的推荐次数

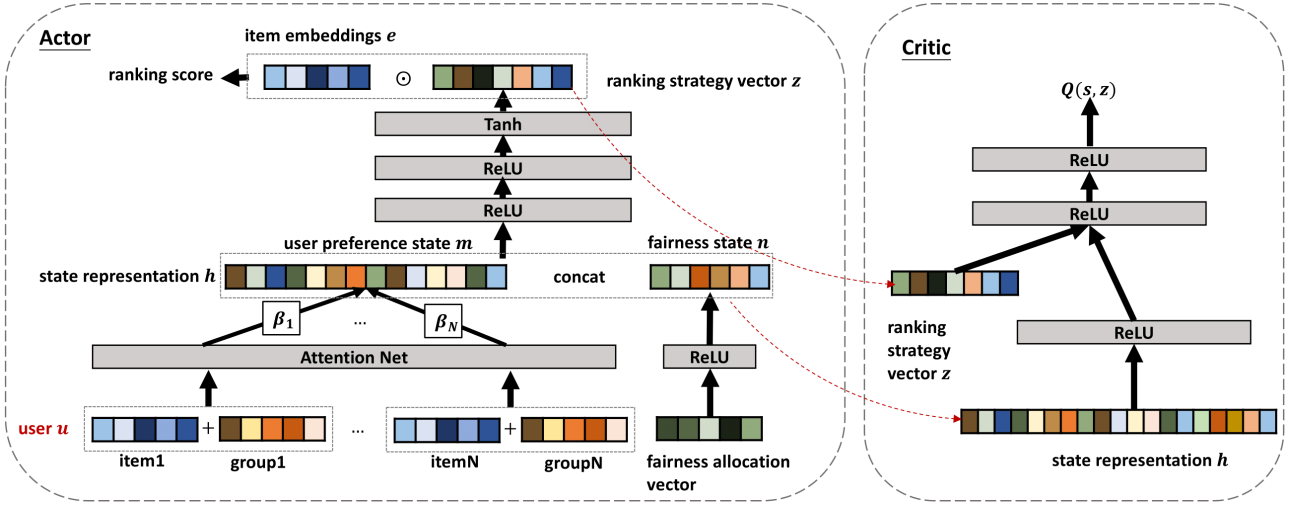
现存方法的问题

- 每次推荐都考虑公平性约束对推荐的精度影响过大。
- 实现公平可以是一个长期的过程，不需要时刻都保证公平
  - 在对包容性比较高的用户进行推荐时注重公平性
  - 在对有特定品味的用户进行推荐时注重准确性
  - 考虑累积公平
- 第一个将推荐的公平性问题建模为一个在线交互(online interactive)过程。



合作项目：使用强化学习进行准确性和公平性的权衡

- 使用强化学习的出发点
  - 用户的兴趣是在动态变化的
  - 不需要时时保证绝对的公平，可以通过对某些对于多样性的推荐结果接受度比较大的时候，侧重公平的推荐，以达到动态平衡。
- 对于用户兴趣和当前系统公平性状态分别建模，并设计相应的激励函数，将二者融合。
- 考虑Disparate Impact这一公平性标准。



合作项目：使用强化学习进行准确性和公平性的权衡

Table 1: Experimental results on MovieLens and Kiva.

	MovieLens			Kiva		
	CVR	PropFair	UFG	CVR	PropFair	UFG
NMF	0.7972	0.8592	4.2362	0.4211	0.8473	1.4635
SVD	0.8478	0.8337	5.4795	0.4870	0.8686	1.6931
DeepFM	0.8612	0.8098	5.8323	0.6349	0.8671	2.3752
LinUCB	0.8577	0.8464	5.9476	0.6517	0.8697	2.4970
DRR	0.8592	0.8470	6.0177	0.6567	0.8645	2.5183
MRPC	0.8361	0.8608	5.2508	0.4286	0.8761	1.5332
FairRec	**0.8702***	**0.8666***	**6.6776***	**0.6905***	**0.8838***	**2.8555***

表1. 所提出的FairRec在不同数据集上的对比实验结果

未来的研究方向

- 如何消除数据中的偏见？
- 考虑哪一方的公平性？商品提供者？商品消费者？二者结合？
- 如何设计合理的公平性指标？短期？长期？
- 如何更好的平衡准确率和公平性？
- ……

THANKS☺

REFERENCES

- [Burke'18] Burke, Robin, et al. "Balanced neighborhoods for multi-sided fairness in recommendation." *FAT*, 2018.
- [Alex'19] Alex Beutel (Google). "Fairness in Recommendation Ranking through Pairwise Comparisons." *KDD*, 2019.
- [Yao'17] Yao, Sirui, and Bert Huang. "Beyond parity: Fairness objectives for collaborative filtering." *NIPS*, 2017.
- [Liu'19] Liu, Weiwen, et al. "Personalized fairness-aware re-ranking for microlending." *RecSys*, 2019.
- [Steck'18] Steck, Harald. "Calibrated recommendations." *RecSys*, 2018.
- [Surer'18] Sürer, Özge, et al. "Multistakeholder recommendation with provider constraints." *RecSys*, 2018.
- [Singh'18] Singh, Ashudeep, and Thorsten Joachims. "Fairness of exposure in rankings." *KDD*, 2018.
- [Biega'18] Biega, Asia J., et al. "Equity of attention: Amortizing individual fairness in rankings." *SIGIR*, 2018.