Towards Long-term Fairness in Recommendation

Presented by Yingqiang Ge

Rutgers University

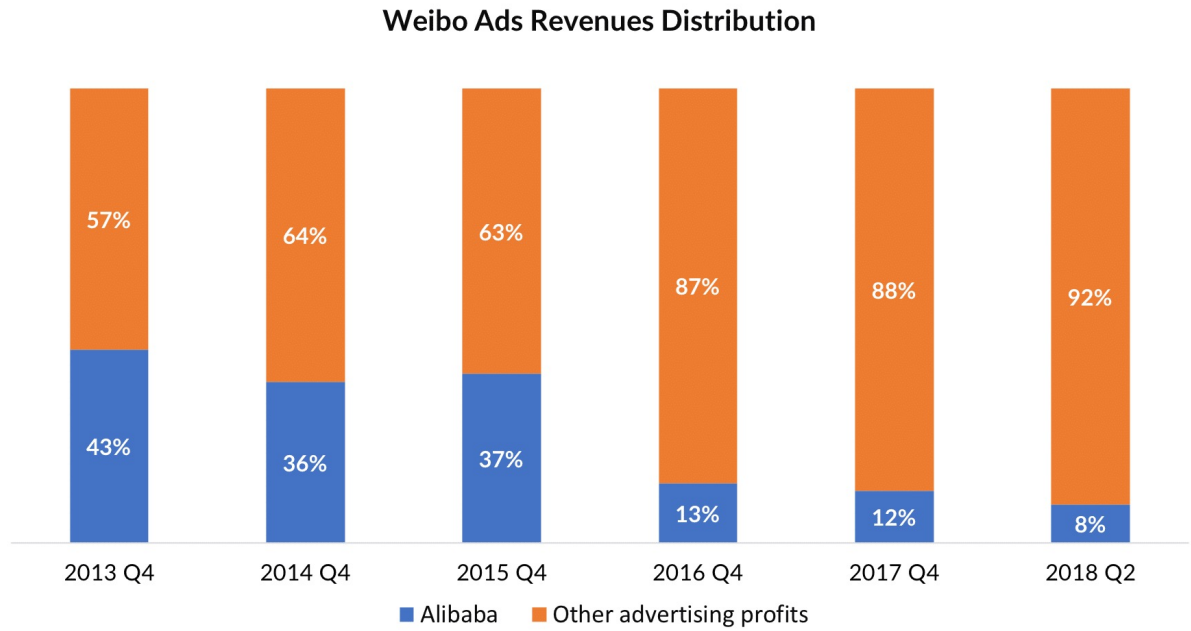
3/3/2021

Agenda

- Motivation
- Problem Formulation
- Model Specification
- Experiments
- Contributions
- Q&A

Motivation

- Why Fairness in E-commerce



Source: Tech.sina.com, Walkthechat Analysis

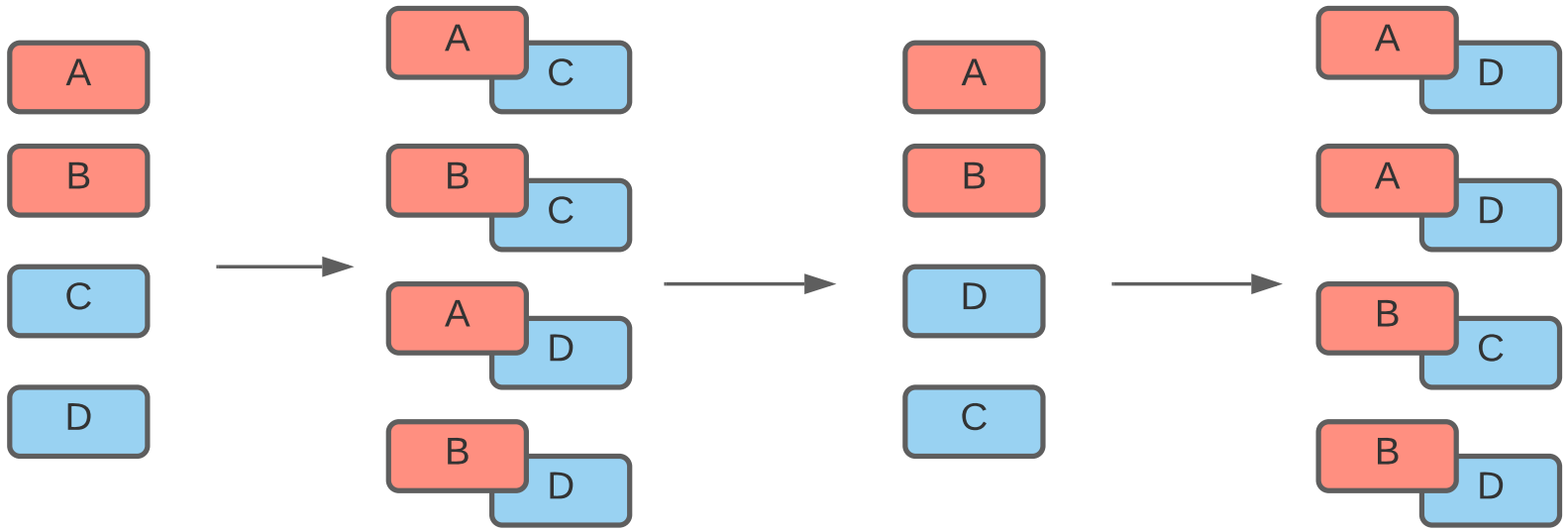
Once the Superstar left the platform(i.e., move to your competitor), the whole ecosystem of the e-commerce platform would crush immediately.

Motivation

- Existing approaches fail to consider the dynamic nature of the recommender systems, where attributes such as item popularity may change over time
- Imposing seemingly fair decisions through static criteria can lead to unexpected unfairness in the long run.
- In essence, fairness cannot be defined in static or one-shot setting without considering the long-term impact, and long-term fairness cannot be achieved without understanding the underlying dynamics.

Motivation

- Consider a simple example of ``Matthew Effect'' in a RS



Motivation

- **Static fairness** does not consider the changes in the recommendation environment.
- **Dynamic fairness** learns a strategy that accommodates the dynamics.
- **Long-term fairness** views the recommendation as a long-term process and aims to maintain fairness in the long run by achieving dynamic fairness over time.

Need a fairness-aware system that

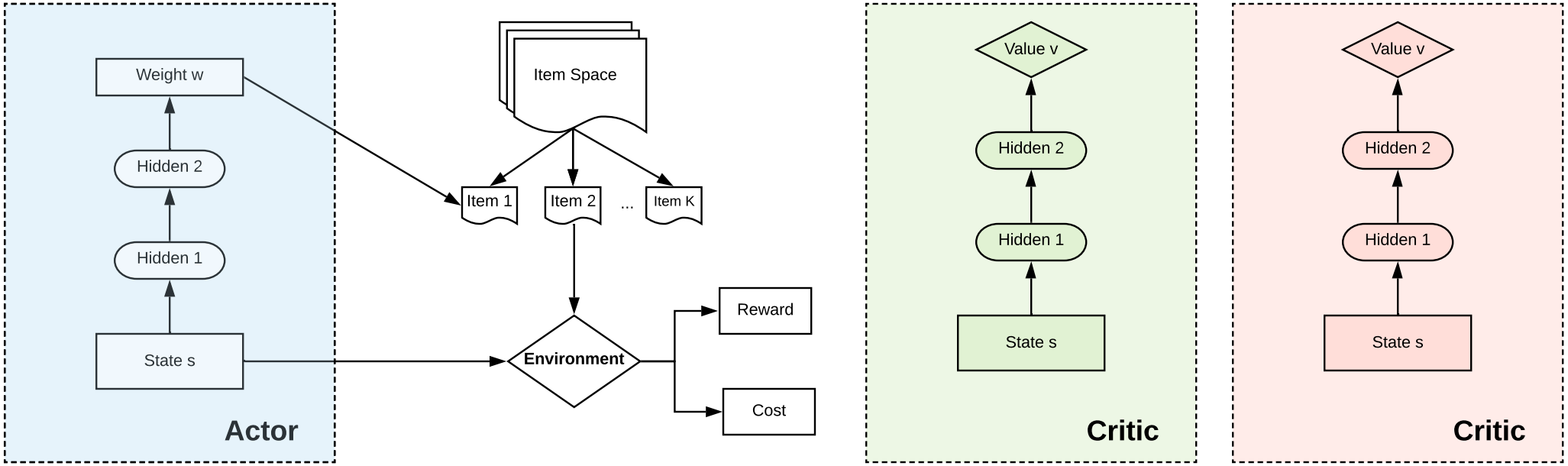
- is dynamic > static.
- considers long-term impact > gives one-time solution.

Problem Formulation

In order to capture the interactive nature of recommendation scenarios, reinforcement learning (RL) based solutions have become an important topic.

- **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=$\{a_t^1, \dots, a_t^K\}$ 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.

Problem Formulation



Problem Formulation

$$\pi_{k+1} = \arg\max_{\pi \in \Pi_\theta} \mathop{\mathrm{E}}_{\substack{s \sim d^{\pi_k} \\ a \sim \pi}} \left[A^{\pi_k}(s,a) \right],$$

$$\text{s.t. } J_{C_i}(\pi_k) + \frac{1}{1-\gamma} \mathop{\mathrm{E}}_{\substack{s \sim d^{\pi_k} \\ a \sim \pi}} \left[A^{\pi_k}_{C_i}(s,a) \right] \leq \mathbf{d}_i, \forall i$$

$$\bar{D}_{KL}(\pi \| \pi_k) \leq \delta$$

Particularly, for problems with only one linear constraint, there is an analytical solution, which is also given by Achiam et al.

$$\theta_{k+1} = \arg\max_{\theta} g^\top (\theta - \theta_k)$$

$$\text{s.t. } c + b^\top (\theta - \theta_k) \leq 0$$

$$\frac{1}{2} (\theta - \theta_k)^\top H (\theta - \theta_k) \leq \delta$$

Fairness Constraints

$$R(s_t, a_t, s_{t+1}) = \sum_{l=1}^{K} \mathbb{1}(a_t^l \text{ gets positive feedback})$$

$$C(s_t, a_t, s_{t+1}) = \sum_{l=1}^{K} \mathbb{1}(a_t^l \textit{ is in sensitive group})$$

Exact-K Fairness Constraint

$$\begin{aligned}
\frac{\text{Exposure}_t\ (G_0)}{\text{Exposure}_t\ (G_1)} &\leq \alpha \\
\text{Exposure}_t\ (G_0) &\leq \alpha \text{Exposure}_t\ (G_1) \\
(1+\alpha)\text{Exposure}_t\ (G_0) &\leq \alpha \text{Exposure}_t\ (G_0) + \alpha \text{Exposure}_t\ (G_1) \\
(1+\alpha)\text{Exposure}_t\ (G_0) &\leq \alpha K \\
C(s_t, a_t, s_{t+1}) &\leq \frac{\alpha}{1+\alpha} K = \alpha' K
\end{aligned}$$

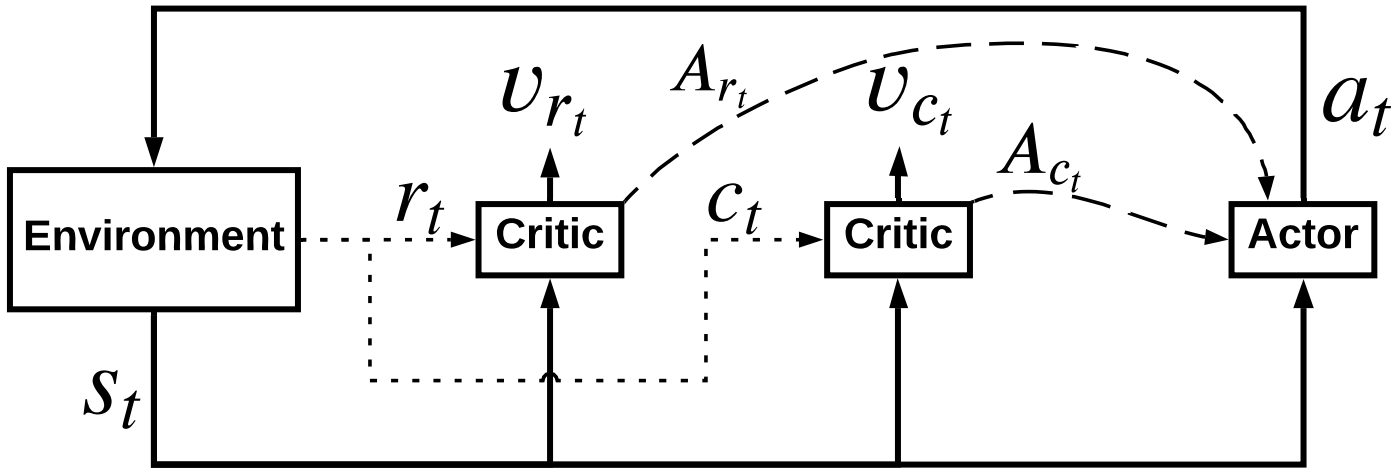
Fairness Constraints

Goal: use CPO to learn a policy that maximize the cumulative reward while the cost is within the bound.

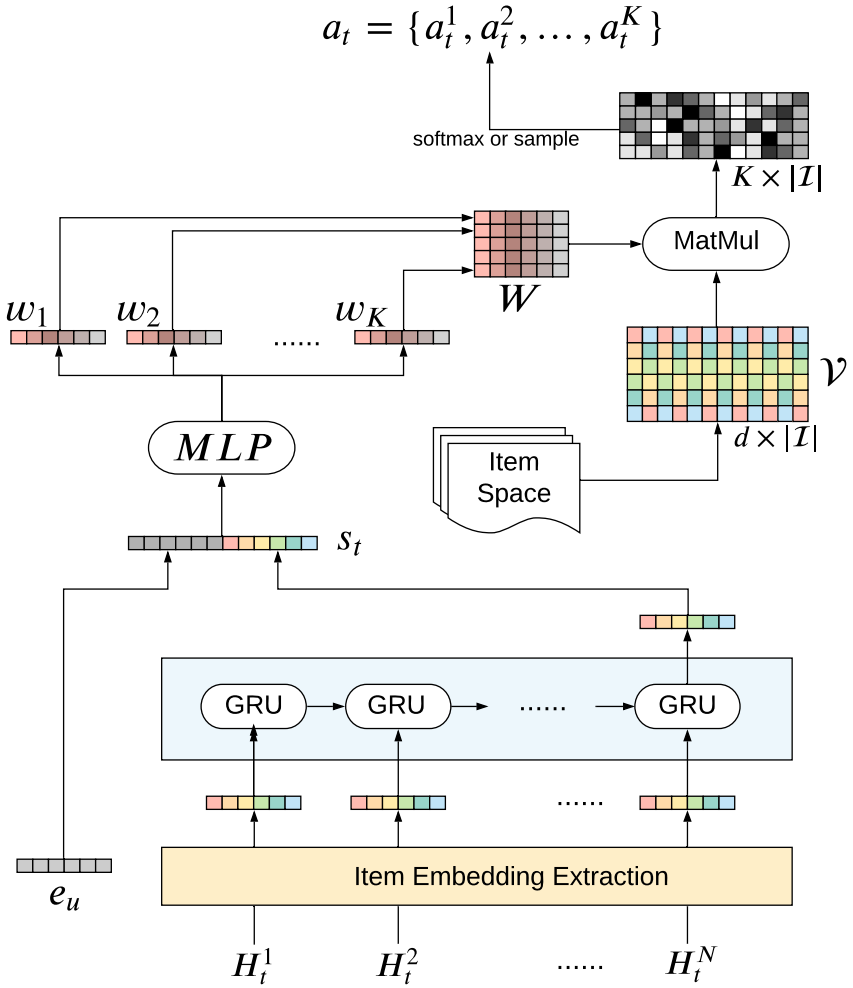
Ratio bound between two groups as in exact-K fairness constraints

$$J_C(\pi) = \mathop{\mathrm{E}}_{\tau \sim \pi} \left[\sum_{t=0}^{T} \gamma_c^T C(s_t, a_t, s_{t+1}) \right] \leq \sum_{t=0}^{T} \gamma_c^t \alpha' K$$

Model Specification



Model Specification



$$s_t = [\mathbf{e}_u; \mathbf{h}_u],\ \mathbf{h}_u = \mathrm{GRU}(H_t)$$

$$H_{t+1} = \begin{cases} \{H_t^2, \ldots, H_t^N, a_t^l\} & r_t^l > 0 \\ H_t & \text{Otherwise} \end{cases}$$

$$P_k = \mathrm{softmax}(W_k \mathcal{V}^\top),\ k = 1, \ldots, K,$$

$$a_t^k = \underset{i \in \{1,\ldots,|\mathcal{I}|\}}{\arg\max}\ P_{k,i},\ \forall k = 1, \ldots, K,$$

Model Specification

Algorithm 1: Parameters Training for FCPO

Input: step size δ, cost limit value $\mathbf{d}$, and line search ratio β
Output: parameters θ, ω and ϕ of actor network, value function, cost function
Randomly initialize θ, ω and ϕ.
Initialize replay buffer D;
for $Round = 1 \dots M$ **do**
    Initialize user state s_0 from log data;
    for $t = 1 \dots T$ **do**
        Observe current state s_t based on Eq. (12);
        Select an action $a_t = \{a_t^1, \dots, a_t^K\} \in \mathcal{I}^K$ based on Eq. (14) and Eq. (15)
        Calculate reward r_t and cost c_t according to environment feedback based on Eq. (8) and Eq. (9);
        Update s_{t+1} based on Eq. (12);
        Store transition $(s_t, a_t, r_t, c_t, s_{t+1})$ in D in its corresponding trajectory.
    end
    Sample minibatch of $\mathcal{N}$ trajectories $\mathcal{T}$ from D;
    Calculate advantage value A, advantage cost value A_c;
    Obtain gradient direction d_θ by solving Eq. (4) with A and A_c;
    repeat
        $\theta' \leftarrow \theta + d_\theta$
        $d_\theta \leftarrow \beta d_\theta$
    until $\pi_{\theta'}(s)$ *in trust region & loss improves & cost* $\leq \mathbf{d}$;
    (Policy update) $\theta \leftarrow \theta'$;
    (Value update) Optimize $\boldsymbol{\omega}$ based on Eq.(16);
    (Cost update) Optimize $\boldsymbol{\phi}$ based on Eq.(17);
end

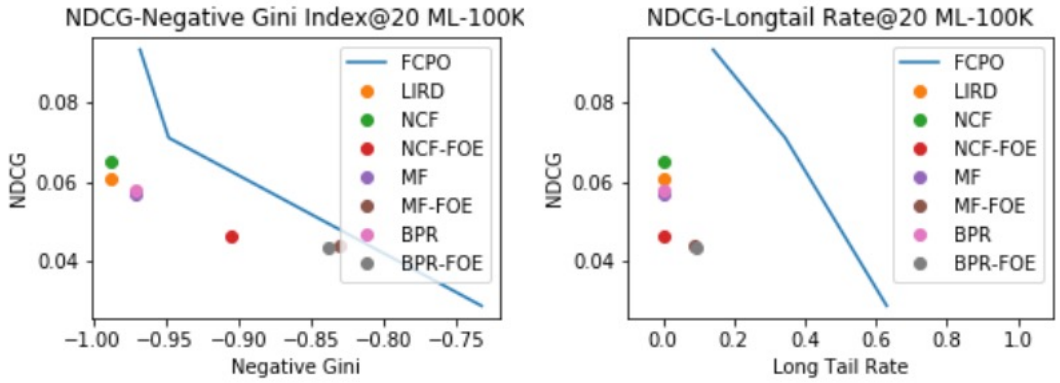
Experiment

Methods	Recall (%) ↑			F1 (%) ↑			NDCG (%) ↑			Gini Index (%) ↓			Popularity Rate (%) ↑		
	K=5	K=10	K=20	K=5	K=10	K=20	K=5	K=10	K=20	K=5	K=10	K=20	K=5	K=10	K=20
Movielens-100K															
MF	1.847	3.785	7.443	2.457	3.780	5.074	3.591	4.240	5.684	98.99	98.37	97.03	99.98	99.96	99.92
BPR-MF	1.304	3.539	8.093	1.824	3.592	5.409	3.025	3.946	5.787	98.74	98.17	97.01	99.87	99.87	99.78
NCF	1.995	3.831	6.983	2.846	4.267	5.383	5.319	5.660	6.510	99.70	99.39	98.80	100.0	100.0	100.0
LIRD	1.769	5.467	8.999	2.199	4.259	4.934	3.025	3.946	5.787	99.70	99.41	98.81	100.0	100.0	100.0
MF-FOE	1.164	2.247	4.179	1.739	2.730	3.794	3.520	3.796	4.367	86.29	84.05	82.98	92.90	91.89	90.98
BPR-FOE	0.974	2.053	4.404	1.496	2.568	3.933	3.127	3.514	4.332	86.50	84.38	83.78	92.17	91.36	90.70
NCF-FOE	1.193	1.987	4.251	1.759	2.398	3.698	4.033	3.897	4.633	96.92	94.53	90.44	100.0	100.0	100.0
FCPO-1	**4.740**	**8.607**	**14.48**	**4.547**	**5.499**	**5.855**	**6.031**	**7.329**	**9.323**	98.73	98.07	96.75	92.60	90.42	85.85
FCPO-2	3.085	5.811	10.41	3.270	4.164	4.953	4.296	5.203	7.104	97.95	96.88	94.78	70.07	68.28	65.55
FCPO-3	0.920	1.668	3.329	1.272	1.807	2.535	2.255	2.369	2.871	**75.23**	**74.06**	**73.23**	**36.52**	**36.66**	**36.94**
Movielens-1M															
MF	1.152	2.352	4.650	1.701	2.814	4.103	3.240	3.686	4.574	99.44	99.18	98.74	99.92	99.90	99.86
BPR-MF	1.240	2.627	5.143	1.773	2.943	4.197	3.078	3.593	4.632	98.93	98.44	97.61	99.40	99.23	98.96
NCF	1.178	2.313	4.589	1.832	2.976	4.382	4.114	4.380	5.080	99.85	99.71	99.42	100.0	100.0	100.0
LIRD	1.961	3.656	5.643	2.673	3.758	4.065	3.078	3.593	4.632	99.87	99.73	99.46	100.0	100.0	95.00
MF-FOE	0.768	1.534	3.220	1.246	2.107	3.345	3.321	3.487	4.021	92.50	91.06	91.32	98.89	98.78	98.68
BPR-FOE	0.860	1.637	3.387	1.374	2.233	3.501	3.389	3.594	4.158	90.48	88.92	89.01	96.56	96.12	95.78
NCF-FOE	0.748	1.403	2.954	1.230	1.980	3.175	3.567	3.589	4.011	97.73	96.57	95.04	100.0	100.0	100.0
FCPO-1	**2.033**	**4.498**	**8.027**	**2.668**	**4.261**	**5.201**	**4.398**	**5.274**	**6.432**	99.81	99.67	99.34	99.28	96.93	91.70
FCPO-2	1.520	3.218	6.417	2.015	3.057	4.145	3.483	3.920	5.133	99.47	99.10	97.41	72.66	68.27	71.35
FCPO-3	0.998	1.925	3.716	1.449	2.185	2.948	2.795	2.987	3.515	**88.97**	**88.34**	**87.70**	**63.43**	**62.73**	**61.45**

Experiment

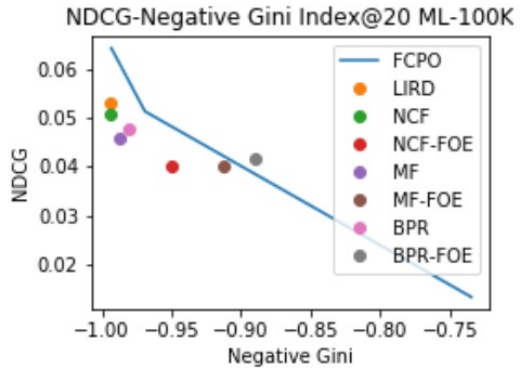
- Our FCPO approach achieves the best top-K recommendation performance against all baselines on both datasets, implying that the proposed method does have the ability to capture dynamic user-item interactions,
- We can easily see that there exists a trade-off between the recommendation performance and the fairness performance both in FCPO and FOE.

Experiment

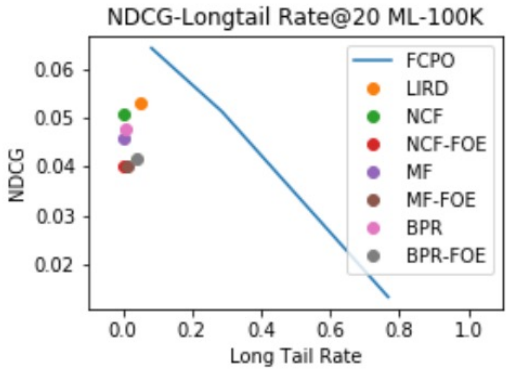


(a) NDCG vs Negative Gini on ML100K

(b) NDCG vs Long-tail Rate on ML100K



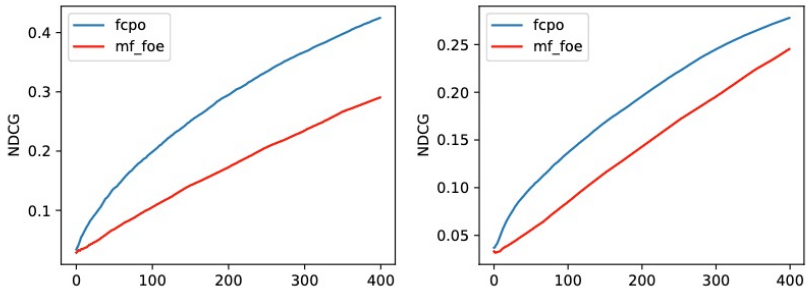
(c) NDCG vs Negative Gini on ML1M



(d) NDCG vs Long-tail Rate on ML1M

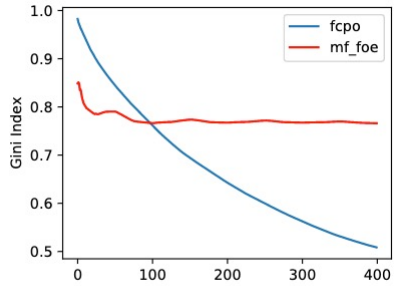
- For the same Gini Index, our method achieves much better NDCG; meanwhile, under the same NDCG scores, our method achieves better fairness.
- FCPO can achieve much better trade-off than FOE in both individual fairness and group fairness.

Experiment

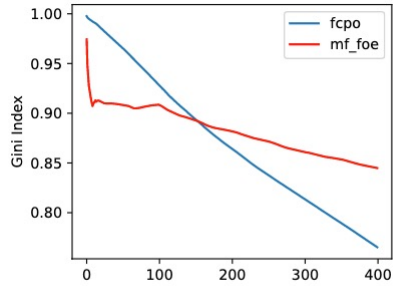


(a) NDCG on Movielens100K

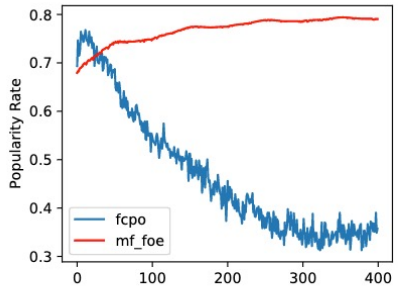
(b) NDCG on Movielens1M



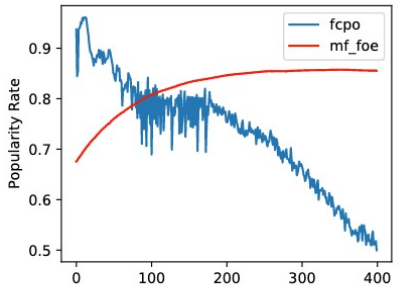
(c) Gini on Movielens100K



(d) Gini on Movielens1M



(e) Popularity Rate on Movielens100K



(f) Popularity Rate on Movielens1M

- We compared FCPO with a static short-term fairness solution (i.e., MF-FOE) for 400 steps of recommendation.
- Since FCPO makes adjustment of its policy according to the fairness feedback, it can successfully and continuously suppress the fairness metric to a much lower value during testing.
- The overall ranking performance of MF-FOE is consistently outperformed by FCPO, which indicates that MF-FOE sacrifices the recommendation performance more than FCPO in order to control fairness.

Contributions

1. We propose to model the sequential interactions between consumers and recommender systems under the fairness constraint of item exposure as a **CMDP**.
2. We leverage the **Constrained Policy Optimization** (CPO) with adapted neural network architecture to automatically learn the optimal policy under different fairness constraints.
3. To the best of our knowledge, this is the first attempt to model the dynamic nature of fairness with respect to **changing group labels**, and to show its effectiveness in the long term.

Q&A

thank you
danke
謝謝
ngiyabonga
rахмат
teşekkür ederim
gracias
merci
obrigado
bedankt
спасибо
Баярлалаа
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
dziękuję
arigatô
dakujem
takk
grazie
tapadh leat
mauruuru
hvala
sagolun
taiku
dank je
mersi
barka
welalin
tack
spas
kia ora
vinaka
спасибі
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
faafetai lava
kiitos
dankie
dhanyavad
nanni
nandri
bayarlalaa
gràçie
köszönöm
dankon
aciu
akun
enkosi
хвала
asante
manana
obrigada
murakoze
tenki
djiere dieuf
tau
дякую
mamnun
chokrane
chnorakaloutioun
gratias ago
gràcies
sulpáy
sobodi
děkuji
mèsi
didi madloba
najis tuke
kam sah hamnida
rahmat
ありがとう
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
mercê
мерси
trugarez
xièxie
감사합니다
ευχαριστώ