Fair Representation Learning for Recommendation: A Mutual Information Perspective (AAAI 2023)

Chen et. al.

Reviewer: Jihu Lee, Jinwon Park

IDEA lab
Department of Statistics
Seoul National University

March, 21, 2024

Table of Contents

Fairness in Recommendation

- Several works exist that deal with fairness in recommendation systems
- While these models successfully mitigate unfair recommendation results to some extent, they still suffered from a substantial drop of recommendation accuracy
- Authors propose a novel two-fold MI based objective from both the user side and item side
- Authors propose the **FairMI** framework for emedding fairness in CF-based recommendations

Table of Contents

Preliminaries

- U: user set ($|U| = M$), V: item set ($|V| = N$)
- $\mathbf{R} \in \mathbb{R}^{M \times N}$: user-item interaction
- r_{uv}: takes 1 when user u has interacted with item i, takes 0 if not
- $\mathcal{G} =< U \cup V, \mathbf{A} >$: user-item bipartite graph

Mutual Information

- Shannon entropy-based measurement for the dependence between two random variable

$$\mathcal{I}\left(\mathbf{X};\mathbf{Y}\right) = \mathcal{H}(\mathbf{X}) - \mathcal{H}\left(\mathbf{X}|\mathbf{Y}\right) \tag{1}$$

Architecture

- 1 sensitive attribute encoder, 1 interest encoder, 2-fold MI based objective
- Basic idea: decompose the embedding $\mathbf{e}$ into a sensitive-aware embedding $\mathbf{e}^{\mathbf{s}}$ and a sensitive-free embedding $\mathbf{e}^{\mathbf{z}}$

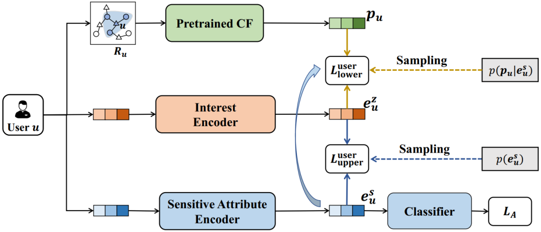

Figure 1: Overall architecture

Sensitive Attribute Encoder

$$
\begin{aligned}
\mathbf{h}_v^{k+1} &= GCN\left(\mathbf{h}_v^k, \left\{\mathbf{h}_u^k : u \in \mathbf{R}_v\right\}\right) \\
\mathbf{h}_u^{k+1} &= GCN\left(\mathbf{h}_u^k, \left\{\mathbf{h}_v^k : v \in \mathbf{R}_u\right\}\right)
\end{aligned} \tag{2}
$$

- $\mathbf{R}_u$ and $\mathbf{R}_v$ denote neighboring nodes of user u and item v
- output: $\mathbf{e}_u^s = \mathbf{h}_u^K, \mathbf{e}_v^s = \mathbf{h}_v^K$
- Apply a sensitive attribute classifier $\mathcal{S}$: $\hat{a}_u = \mathcal{S}(\mathbf{e}_u^s)$

$$
\min_{\theta_{\mathcal{S}}, \mathbf{E}^s} \mathcal{L}_A = -\frac{1}{M} \sum_{u=1}^{M} a_u \log\left(\hat{a}_u\right) \tag{3}
$$

Interest Encoder

User condition

1. Sensitive-free user embedding $\mathbf{e}_u^z$ should have no MI with sensitive-aware user embedding $\mathbf{e}_u^s$
2. Sensitive-free user embedding $\mathbf{e}_u^z$ should have maximum MI with user interactions $\mathbf{R}_u$, conditioned on sesitive-aware user embedding $\mathbf{e}_u^s$

Item condition

3. Sensitive-free item embedding $\mathbf{e}_v^z$ should have no MI with sensitive-aware item embedding $\mathbf{e}_v^s$
4. Sensitive-free item embedding $\mathbf{e}_v^z$ should have maximum MI with user interactions $\mathbf{R}_v$, conditioned on sesitive-aware item embedding $\mathbf{e}_v^s$

- Condition 1&3 $\rightarrow$ minimize $\mathcal{I}\left(\mathbf{e}_u^z; \mathbf{e}_u^s\right)$ and $\mathcal{I}\left(\mathbf{e}_v^z; \mathbf{e}_v^s\right)$
- Condition 2&4 $\rightarrow$ maximize $\mathcal{I}\left(\mathbf{e}_u^z; \mathbf{R}_u | \mathbf{e}_u^s\right)$ and $\mathcal{I}\left(\mathbf{e}_v^z; \mathbf{R}_v | \mathbf{e}_v^s\right)$

Overall loss

$$\min_{\mathbf{E}^z} \mathcal{L}_{\mathsf{all}} = \mathcal{L}_{\mathsf{rec}} + \mathcal{L}_{\mathsf{MI}} \tag{4}$$

where $\mathcal{L}_{\mathsf{rec}}$ can be any recommendation loss (e.g. BPR loss)

MI Upper Bound

Proposition 1

Given $\mathbf{e}_j^s \sim p(\mathbf{e}_u^s)$*, if the conditional distribution* $p\left(\mathbf{e}_u^s | \mathbf{e}_u^z\right)$ *is known, then*

$$\mathcal{I}\left(\mathbf{e}_u^s; \mathbf{e}_u^z\right) \leq \mathbb{E}\left[\log p\left(\mathbf{e}_u^s | \mathbf{e}_u^z\right) - \frac{1}{M}\sum_{j=1}^{M} \log p\left(\mathbf{e}_j^s | \mathbf{e}_u^z\right)\right] \tag{5}$$

$$\min_{q_\phi} \mathbb{D}_{\mathsf{KL}}\left[q_\phi(\mathbf{e}_u^s | \mathbf{e}_u^z) || p(\mathbf{e}_u^s | \mathbf{e}_u^z)\right] \tag{6}$$

$$\begin{aligned} &\min_{\mathbf{e}_u^z} \mathcal{L}_{\text{upper}}^{\text{user}} \\ &= \frac{1}{M}\sum_{u=1}^{M}\left[\log q_\phi\left(\mathbf{e}_u^s | \mathbf{e}_u^z\right) - \frac{1}{M}\sum_{j=1}^{M} \log q_\phi\left(\mathbf{e}_j^s | \mathbf{e}_u^z\right)\right] \end{aligned} \tag{7}$$

MI Lower bound

Due to the high-dimension and sparsity of the user historical interactions, authors leverage a pre-trained models (e.g., BPR, LightGCN) to generate low-rank embedding $\mathbf{p}_u$ to denote $\mathbf{R}_u$.

Proposition 2

Given $\mathbf{p}_u, \mathbf{e}_u^z, \mathbf{e}_u^s \sim p(\cdot, \cdot\cdot)$, $\mathbf{p}_j \sim p(\mathbf{p}_u | \mathbf{e}_u^s)$, *with a score function* f, *we have*

$$\mathcal{I}(\mathbf{e}_u^z; \mathbf{p}_u | \mathbf{e}_u^s) \leq \mathbb{E}\left[\log \frac{\exp f(\mathbf{p}_u, \mathbf{e}_u^z, \mathbf{e}_u^s)}{\frac{1}{M}\sum_{j=1}^{M} \exp f(\mathbf{p}_j, \mathbf{e}_u^z, \mathbf{e}_u^s)}\right] \tag{8}$$

$$\begin{aligned} &\max_{\mathbf{e}_u^z} \mathcal{L}_{\text{lower}}^{\text{user}} \\ &= \frac{1}{M}\sum_{u=1}^{M}\left[\log \frac{\exp(\text{sim}(\mathbf{p}_u, w(\mathbf{e}_u^z, \mathbf{e}_u^s, \alpha)))}{\frac{1}{M}\sum_{j=1}^{M}\exp(\text{sim}(\mathbf{p}_j, w(\mathbf{e}_u^z, \mathbf{e}_u^s, \alpha)))}\right] \end{aligned} \tag{9}$$

where $w(\mathbf{e}_u^z, \mathbf{e}_u^s, \alpha) = \mathbf{e}_u^z + \alpha \cdot \mathbf{e}_u^s$. ($f$: weighted cosine similarity)

MI bounds

Two-fold MI based loss

$$\mathcal{L}_{\text{MI}} = \beta\left(\mathcal{L}_{\text{upper}}^{\text{user}} + \mathcal{L}_{\text{upper}}^{\text{item}}\right) - \gamma\left(\mathcal{L}_{\text{lower}}^{\text{user}} + \mathcal{L}_{\text{lower}}^{\text{item}}\right) \tag{10}$$

Table of Contents

Datasets

- MovieLens-1M
- Lastfm-360K
- Sensitive attribute: **gender**

Replacement of DP

$$\forall v \in V, f^v_{G_0} = \frac{\sum_{u \in G_0} \mathbf{I}_{v \in TopK_u}}{|G_0|}, f^v_{G_1} = \frac{\sum_{u \in G_1} \mathbf{I}_{v \in TopK_u}}{|G_1|}$$
$$\mathbf{f}_{G_0} = \left[f^1_{G_0}, \ldots, f^v_{G_0}, \ldots, f^N_{G_0}\right], \mathbf{f}_{G_1} = \left[f^1_{G_1}, \ldots, f^v_{G_1}, \ldots, f^N_{G_1}\right] \quad (11)$$

- G_0, G_1: user group with different sensitive
- $TopK_u$: Top-K ranked items for user u

$$DP@K = JSD(\mathbf{f}_{G_0}, \mathbf{f}_{G_1}) \quad (12)$$

Replacement of EO

similar

Accuracy and Fairness Performance

Model	K	NDCG@K↑		RECALL@K↑		DP@K↓		EO@K↓	
		10	20	10	20	10	20	10	20
BPR	Base	0.1943	0.2537	0.1437	0.2280	0.2854	0.2572	0.3580	0.3316
	DP	0.1899	0.2490	0.1409	0.2240	0.2187	0.1870	0.3231	0.2944
	Adv	0.1900	0.2485	0.1404	0.2230	0.1684	0.1363	0.2736	0.2499
	FairRec	0.1896	0.2485	0.1407	0.2236	0.1656	0.1317	0.2714	0.2451
	FairMI*	**0.2022**	**0.2607**	0.1487	**0.2326**	0.1501	0.1285	0.2406	0.2161
	FairMI	**0.2022**	0.2606	**0.1491**	0.2324	**0.1381**	**0.1179**	**0.2233**	**0.2038**
GCN	Base	0.2025	0.2671	0.1523	0.2449	0.2937	0.2626	0.3621	0.3325
	DP	0.1981	0.2603	0.1481	0.2363	0.2297	0.1924	0.3247	0.2955
	Adv	0.1970	0.2579	0.1474	0.2346	0.1517	0.1183	0.2646	0.2338
	FairRec	0.1950	0.2561	0.1472	0.2339	0.1536	0.1193	0.2590	0.2283
	FairGo	0.1822	0.2373	0.1336	0.2108	0.2728	0.2436	0.3382	0.3101
	FairGNN	0.1964	0.2569	0.1466	0.2323	0.1472	0.1181	0.2608	0.2320
	FairMI*	**0.2128**	**0.2754**	0.1581	0.2473	0.1597	0.1340	0.2426	0.2243
	FairMI	**0.2128**	0.2752	**0.1586**	**0.2477**	**0.1337**	**0.1111**	**0.2228**	**0.2006**

Figure 2: MovieLens-1M

Model Analysis

- (Ablation study) Effectiveness of Lower bound and Upper bound
- (Parameter sensitivity analysis) different β and γ


THANK
you!