Bias Mitigation in Recommender Systems to Improve Diversity

Du Cheng[1], Doruk Kilitcioglu[1], & Serdar Kadıoğlu[1,2]

[1] AI Center of Excellence, Fidelity Investments
[2] Computer Science Department, Brown University

Introduction

Bias Mitigation in Recommender Systems to Improve Diversity

Evaluating Recommender Systems is not a trivial or straightforward process.

Optimizing for metrics like Hit-Rate may lead to higher degree of bias among protected groups

We introduce two bias mitigation methods aimed at improving recommendation fairness

We show these methods improve the fairness of recommendations across a protected class

Alternating Least Squares

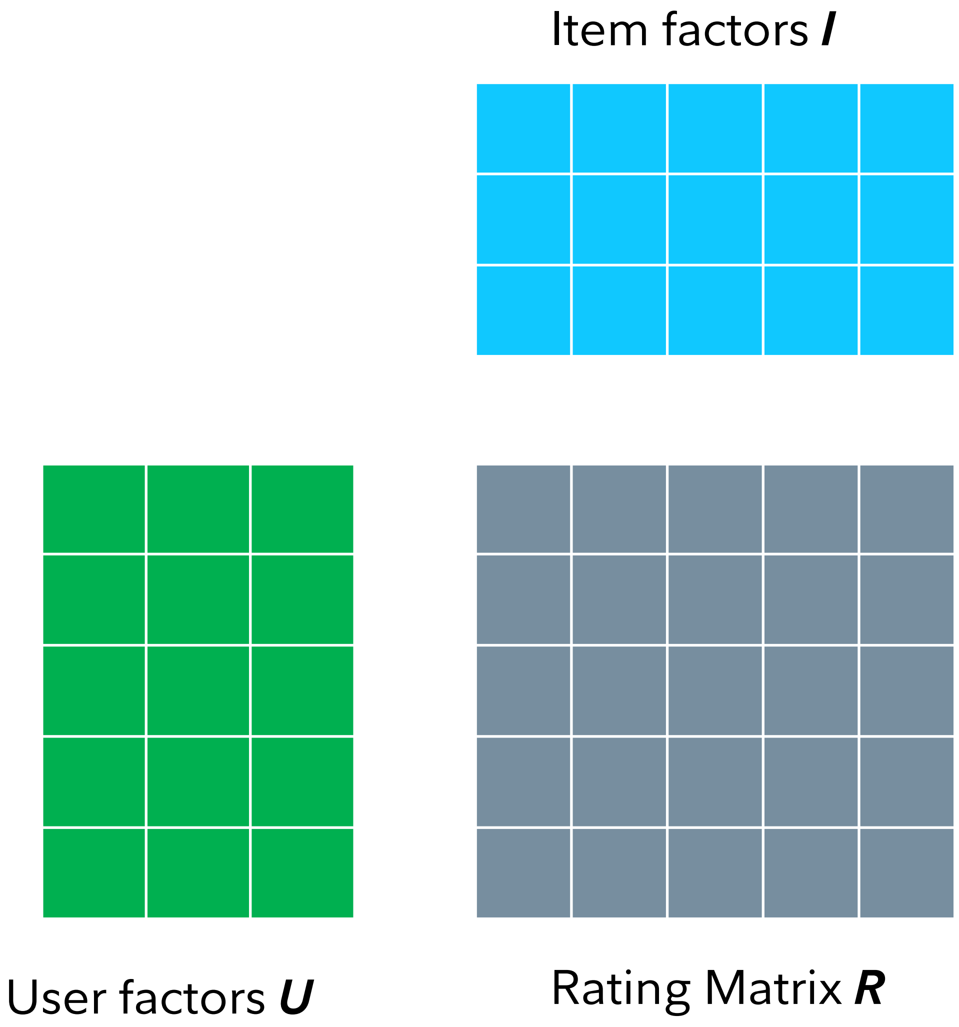
Matrix Factorization Method

- Using user-item interaction matrix ***R***, find user factors ***U*** and item factors ***I*** such that ***U*I* ≈ *R***

- **Optimize** the cost function:

$$\min_{x,y} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda\left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2\right)$$

- **Alternate** between optimizing user factors & item factors



Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: 2008 Eighth IEEE International Conference on Data Mining, 2008, pp. 263–272.

Averaging ALS

Bias Mitigation in Recommender Systems to Improve Diversity

- Train and Finetune Baseline ALS Model
- Slice user data by user_activity bins [1, 10, 1000] and train separate models on different user groups
- Obtain recommendations and scores from averaging the baseline model and user-activity model

Baseline ALS Model

user_id	item_1	item_2	item_3
1	0.6	0.5	0.8
…	0.7	0.4	0.9
2224	0.7	0.9	0.9
2225	0.8	0.9	0.8
…	0.5	0.4	0.3
23358	0.6	0.8	0.7
23359	0.8	0.7	0.9
23360	0.8	0.5	0.8
…	…	…	…
29733	0.9	0.6	0.8

\+

user_id	item_1	item_2	item_3
1	0.5	0.8	
…	…	…	
2224	0.8	0.6	
2225	0.6		0.4
…	…		…
23358	0.3		0.6
23359	0.8	0.9	0.6
23360	0.9	0.8	0.9
…	…	…	…
29733	0.9	0.7	0.6

Low activity

Mid Activity

High Activity

Post-processing

Bias Mitigation

We apply Equalized Odds per item, using **user activity** as a protected class.

Equalized Odds aims to nudge the likelihoods such that the following is true:

$$\Pr\left(\hat{Y} = 1 | A = 0, Y = y\right) = \Pr\left(\hat{Y} = 1 | A = 1, Y = y\right), y \in (0,1)$$

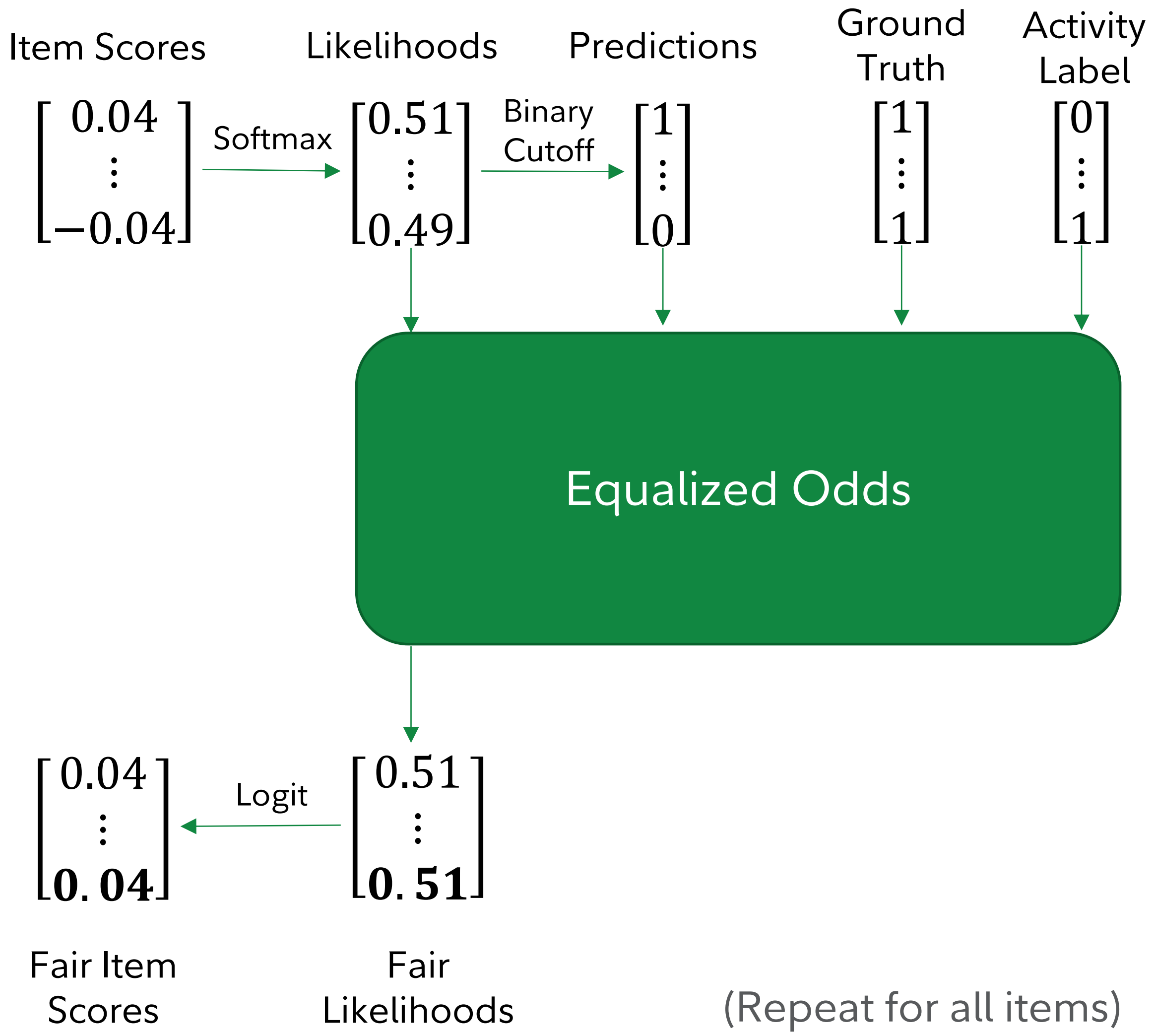
where $\hat{Y}$ is the predicted class, A is binarized user activity, Y is actual class

In other words, we want each item to be **unbiased** when it comes to users with different activity levels. We want to **mitigate** the impact the user's activity has on an item being recommended.

Post-processing

Algorithm

1. Obtain the user-item-score matrix for user-item pairs and run softmax on it.
2. Calculate a binary cutoff point per item based on the 80% quantile scores.
3. Binarize the results item-wise and run equalized odds, using user activity as a protected class.
4. In case the application of equalized odds change the binary label, go back to the softmax scores and use its complement, i.e., (1 - softmax).
5. Re-order recommendations using the new scores



$$\begin{bmatrix} 0.04 \\ \vdots \\ -0.04 \end{bmatrix} \xrightarrow{\text{Softmax}} \begin{bmatrix} 0.51 \\ \vdots \\ 0.49 \end{bmatrix} \xrightarrow{\text{Binary Cutoff}} \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} \quad \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \begin{bmatrix} 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.04 \\ \vdots \\ \mathbf{0.04} \end{bmatrix} \xleftarrow{\text{Logit}} \begin{bmatrix} 0.51 \\ \vdots \\ \mathbf{0.51} \end{bmatrix}$$

Results

Bias Mitigation in Recommender Systems to Improve Diversity

Model	Score	Hit Rate	MRED_USER_ACTIVITY	Runtime
CBOW Baseline	-1.212	0.036	-0.022	N/A
ALS	-21.823	0.046	-0.007	3 min per fold
Separate ALS	-100	0.004	-0.001	10 min per fold
ALS + Averaging (with n_sum = 500)	-11.31	0.027	-0.0086	19 min per fold
ALS + Averaging (with n_sum = 1000)	-6.670	0.017	-0.005	25 min per fold
ALS + Post-processing	-18.761	0.042	-0.006	4 min per fold

Summary

- We focused our work on MRED_USER_ACTIVITY
- Post-processing and Averaging scores from baseline ALS/separate ALS models provide balance between hit rate and diversity metrics

Next Steps

Bias Mitigation in Recommender Systems to Improve Diversity

1. Combine post-processing steps with the Averaging model

2. Extend post-processing to multi-group bias mitigation, such that fairness is improved beyond a single protected group

Fidelity
INVESTMENTS