Popularity Bias in Dynamic Recommendation

Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee
Texas A&M University
Popular items are overly exposed in recommendations at the expense of less popular items that users may find interesting being under-recommended.
Prior works study the popularity bias in a **static** setting.
Prior works study the popularity bias in a **static** setting.
Research gap: popularity bias in *dynamic* recommendation

How does the popularity bias *evolve* in a real-world *dynamic recommendation* process?
Dynamic recommendation
Dynamic recommendation
Dynamic recommendation
Dynamic recommendation
Dynamic recommendation
Key factor: inherent audience size imbalance
Key factor: inherent audience size imbalance

**Inherent Audience Size**
(how many users will click the item if recommended to all users)

<table>
<thead>
<tr>
<th>Item</th>
<th>Audience Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dress</td>
<td>🍒🍒🍒🍒🍒</td>
</tr>
<tr>
<td>Shirt</td>
<td>🍒🍒🍒🍒</td>
</tr>
<tr>
<td>Boot</td>
<td>🍒🍒🍒</td>
</tr>
<tr>
<td>Shoe</td>
<td>🍒🍒</td>
</tr>
<tr>
<td>Sneakers</td>
<td>🍒</td>
</tr>
</tbody>
</table>

A few items may have very large audience sizes (liked by most of users in ground truth), while the majority have small ones.
Key factor: model bias
Key factor: model bias

The recommendation model itself may amplify any imbalances in the data it ingests for training.
Key factor: position bias
Key factor: position bias

Once the model makes recommendations, the top-ranked items are more likely to be examined by users.
Key factor: closed feedback loop
Key factor: closed feedback loop

Bias transfers through the feedback loop

As the closed loop repeating, the feedback data collected from recommendations made by the current model will impact the training of future versions of the model.
Contributions

• Conduct a comprehensive **empirical study** by simulation experiments to investigate how the popularity bias **evolves** in dynamic recommendation, and how the **four factors impact** the bias;

• Proposed a simple but powerful **dynamic debiasing framework** to adapt exiting static debiasing methods to the dynamic scenario;

• Report on extensive experiments to show the **effectiveness** of the proposed dynamic debiasing method compared with the existing static methods.
Outline

• Motivation and Introduction

➢ Problem Formalizations
  • Formalize the dynamic recommendation process
  • Formalize the popularity bias

• Data-driven Study

• Debiasing and Experiment
Algorithm 1: Dynamic Recommendation Process

1. **Bootstrap**: Randomly show $K$ items to each user and collect initial clicks $\mathcal{D}$ and train the first model $\psi$ by $\mathcal{D}$;
Problem formalization: dynamic recommendation process

---

**Algorithm 1:** Dynamic Recommendation Process

1. **Bootstrap:** Randomly show $K$ items to each user and collect initial clicks $\mathcal{D}$ and train the first model $\psi$ by $\mathcal{D}$;

2. **for** $t = 1 : T$ **do**

3. 

4. 

5. 

6. 

---
Problem formalization: dynamic recommendation process

Algorithm 1: Dynamic Recommendation Process

1. **Bootstrap**: Randomly show $K$ items to each user and collect initial clicks $\mathcal{D}$ and train the first model $\psi$ by $\mathcal{D}$;

2. **for** $t = 1 : T$ **do**
   
   3. Recommend $K$ items to the current user $u_t$ by $\psi$;

   4.

   5.

   6.
Problem formalization: dynamic recommendation process

Algorithm 1: Dynamic Recommendation Process

1. **Bootstrap**: Randomly show $K$ items to each user and collect initial clicks $\mathcal{D}$ and train the first model $\psi$ by $\mathcal{D}$;

2. **for** $t = 1 : T$ **do**
   3. Recommend $K$ items to the current user $u_t$ by $\psi$;
   4. Collect new clicks and add them to $\mathcal{D}$;
Problem formalization: dynamic recommendation process

\textbf{Algorithm 1:} Dynamic Recommendation Process

1. **Bootstrap:** Randomly show $K$ items to each user and collect initial clicks $\mathcal{D}$ and train the first model $\psi$ by $\mathcal{D}$;

2. \textbf{for} $t = 1 : T$ \textbf{do}

3. \hspace{1em} Recommend $K$ items to the current user $u_t$ by $\psi$;

4. \hspace{1em} Collect new clicks and add them to $\mathcal{D}$;

5. \hspace{1em} \textbf{if} $t \% L = 0$ \textbf{then}

6. \hspace{2em} Retrain $\psi$ by $\mathcal{D}$;
Problem formalization: popularity bias

Compared with less popular items, whether popular items are more likely to be correctly recommended to matched users who like them?
Problem formalization: popularity bias

- During testing, calculate the **average exposure** every item gets to their **matched users**.
- Sort items based on their popularity.
- Calculate **Gini Coefficient** of exposure over sorted items to evaluate the popularity bias. (higher Gini Coefficient, more severe bias)
Outline

• Motivation and Introduction
• Problem Formalizations
  ➢ Data-driven Study
    • Evolution of the popularity bias
    • Impacts of the four bias factors
• Debiasing and Experiment
Data-driven study: evolution of popularity bias

- Position bias
- Closed feedback loop
- Model bias
- Inherent audience size imbalance

![Graph showing Gini Coefficient over iterations for Popular, Random, and MF categories. The graph indicates the evolution of popularity bias over iterations, withPopular and MF showing a stabilization of the Gini Coefficient, while Random remains at a lower level.](image-url)
Data-driven study: evolution of popularity bias

- Position bias
- Closed feedback loop
- Model bias
- Inherent audience size imbalance

Increases rapidly then keeps at a high level
Data-driven study: impact of position bias

- Position bias
- Closed feedback loop
- Model bias
- Inherent audience size imbalance

Graph showing Gini Coefficient over iterations with and without position bias.
Data-driven study: impact of position bias

✗ Position bias
✓ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

By counteracting the position bias, lower bias is achieved
Data-driven study: impact of position bias

✗ Position bias
✓ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

But the overall pattern of the bias is still the same
Data-driven study: impact of position bias

✓ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

Conclusion: Position bias can intensify the popularity bias.
Data-driven study: impact of closed feedback loop

- Position bias
- Closed feedback loop
- Model bias
- Inherent audience size imbalance

Graph showing the Gini coefficient over iterations with and without a closed feedback loop.
Data-driven study: impact of closed feedback loop

✗ Position bias
✗ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

By breaking the closed feedback loop, bias increases slower
Data-driven study: impact of closed feedback loop

✗ Position bias
✗ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

But the bias keeps increasing and reach high level
Data-driven study: impact of closed feedback loop

✗ Position bias
✗ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

Conclusion: Closed feedback loop can intensify the popularity bias.
Data-driven study: impact of closed feedback loop

✗ Position bias
✗ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

Conclusion: Model bias and inherent audience size imbalance are the main source of popularity bias,
Data-driven study: impact of model bias

✗ Position bias
✗ Closed feedback loop
✓ Model bias
✓ Inherent audience size imbalance

more skewed training data

more dense training data
Data-driven study: conclusions

• Inherent audience size imbalance and model bias are the main sources of popularity bias, which can produce the bias without existence of other factors;

• Position bias and closed feedback loop can intensify the bias when inherent audience size imbalance and model bias exist;

• Higher training data density and greater imbalance can increase the effect of model bias.
Outline

• Motivation and Introduction
• Problem Formalizations
• Data-driven Study

➤ Debiasing and Experiment
Debias in a dynamic way

1. Adopt an **existing static debiasing method**, apply it to dynamic recommendation process by gradually **increasing debiasing strength**;

2. Correct predicted scores based on **false positive signals**.
Debias in a dynamic way

• Adopt an existing static debiasing method, apply it to dynamic recommendation process by gradually increasing debiasing strength;
Debias in a dynamic way

- Adopt an **existing static debiasing method**, apply it to dynamic recommendation process by gradually **increasing debiasing strength**;

Example: an existing debiasing method **Scale**

\[
\hat{r}_{u,i}^{(scaled)} = \frac{\hat{r}_{u,i}^{(model)}}{(C_i)^\alpha}
\]

Debiasing strength hyper-parameter

popularity of item \(i\)
Debias in a dynamic way

• Adopt an **existing static debiasing method**, apply it to dynamic recommendation process by gradually **increasing debiasing strength**;

Example: an existing debiasing method **Scale**

\[
\hat{r}_{u,i}^{(scaled)} = \frac{\hat{r}_{u,i}^{(model)}}{(C_i)^{\alpha}}
\]
Debias in a dynamic way

- Correct predicted scores based on **false positive signals**.

False positive signals: the cases that some items are recommended to some users but receive no feedback
Debias in a dynamic way

• Correct predicted scores based on **false positive signals**.

The probability user \( u \) likes item \( i \) given \( i \) has been recommended to \( u \) for \( F \) times and did not receive any clicks.

\[
P(r_{u,i} = 1 | c_{k_1} = 0, \ldots, c_{k_F} = 0)
\]

The \( F \text{th} \) time item \( i \) being recommended to user \( u \), ranked at \( k_F \) position, and no click happened

The probability user \( u \) likes item \( i \) given \( i \) has been recommended to \( u \) for \( F \) times and did not receive any clicks.
Debias in a dynamic way

• Correct predicted scores based on false positive signals.

\[ P(r_{u,i} = 1 | c_{k_1} = 0, \ldots, c_{k_F} = 0) \]

\[ = 1 - \frac{1 - \theta_{u,i}}{\prod_{f=1}^{F} (1 - \delta_{k_f} \theta_{u,i})} \]

User-item relevance probability, use the prediction from a recommendation model, such as \( \hat{r}_{u,i}^{(scaled)} \) from the previous step.

Examine probability at \( k_f \) position (same as the position bias)
Debiasing experiments

- With increasing debiasing strength, we can **continuously decrease** the bias.
- Fix the debiasing strength as static debiasing method, the bias starts low but **grows** to high level.
Debiasing experiments

MF: 67,816
Scale: 66,630
DScale: 68,645
Debiasing experiments

• Integrate DScale and false positive correction, the popularity **bias is further decreased**;
• **More clicks** are collected by debiasing by the proposed method (higher recommendation utility is achieved).
Debiasing experiments

MF: 67,816
DScale: 68,645
FPC-DScale: 73,145
Debiasing experiments

More experimental details and results can be found in the paper, including:

• Detailed experiment **setup**;
• Experiments on other **datasets** of different levels of skewness;
• Experiments with other **baseline** debiasing methods;
Conclusions

• Conduct a comprehensive empirical study by simulation experiments to investigate how the popularity bias evolves in dynamic recommendation, and how the four factors impact the bias;

• Proposed a simple but powerful dynamic debiasing framework to adapt exiting static debiasing methods to the dynamic scenario;

• Report on extensive experiments to show the effectiveness of the proposed dynamic debiasing method to debias and also increase the recommendation utility.
Thank You!

Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee
Texas A&M University