Combating Heterogeneous Model Biases in Recommendations via Boosting

Jinhao Pan[1], James Caverlee[2] and Ziwei Zhu[1]

[1]Department of Computer Science, George Mason University, USA

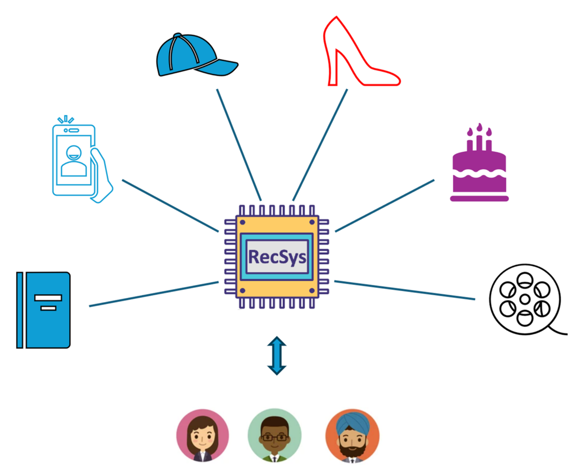
[2]Department of Computer Science and Engineering, Texas A&M University, USA

{jpan23, zzhu20}@gmu.edu, caverlee@tamu.edu

Introduction

Collaborative Filtering (CF) based recommenders often exhibit model biases. Prior research approaches these biases as **isolated** and **standalone** issues, ignoring their interconnected nature and developing separate methods, thereby compromising the specialized debiasing efforts. our goal is to create a holistic debiasing framework capable of **addressing heterogeneous model biases simultaneously**.

Are different users served equally by RecSys?
Are different items recommended equally by RecSys?



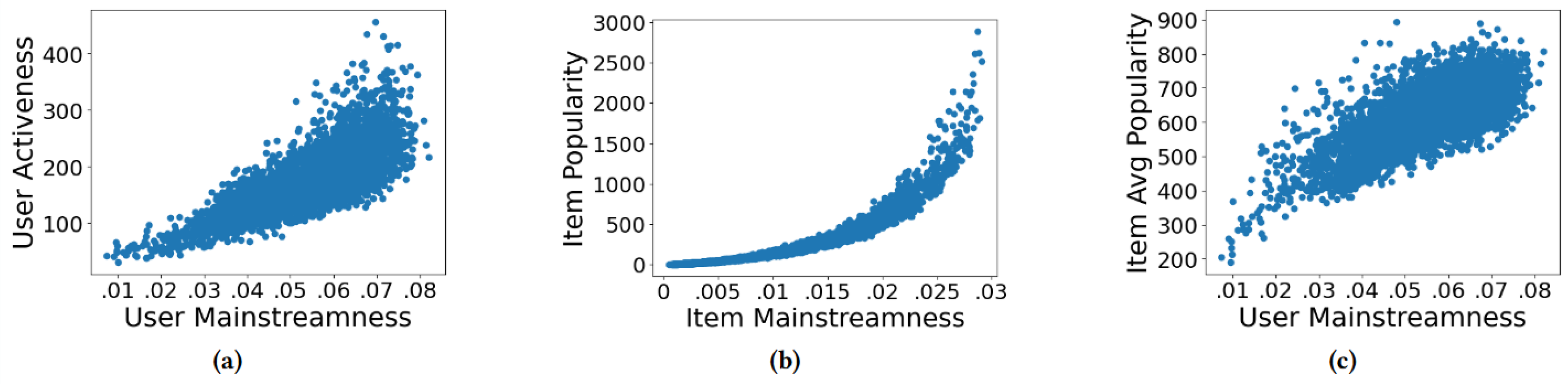
- **Contributions:**

(i) Propose a boosting-based collaborative filtering framework (**CFBoost**) that tailors sub-models for each user and item to mitigate heterogeneous model biases.

(ii) Our theoretical analysis demonstrates that the model achieves an **exponentially decreasing upper bound** on the training loss for any user-item pairs with increasing boosting iterations.

(iii) Extensive **experiments** show that the CFBoost outperforms state-of-the-art debiasing baselines.

Heterogenous Model Biases & Interconnections



(a) (b) (c)

- For four bias dimensions, we sort users/items in a non-descending order by the corresponding score (user mainstreamness/use activeness/item mainstreamness/item popularity) and evenly split them into 5 subgroups to evaluate each bias.

- **Goal: Improve** recommendation utility for niche/inactive users and niche/unpopular items while **preserving** utility for mainstream/active users and mainstream/popular items in a holistic method.

- **User Mainstream Bias:** Niche users receive poor recommendation services.
- **User Activeness Bias:** Inactive users receive poor recommendation services.
- **Item Mainstream Bias:** Niche items are poorly recommended to audiences
- **Item Popularity Bias:** Niche items are poorly recommended to audiences.

Proposed Model

- **Algorithm:**

Algorithm 1: CFBoost

Input : $\mathcal{O} = \{(u, i)\}$; number of boosting iterations $\mathcal{T}$.

Output: Final Prediction: $\widehat{\mathcal{O}}$

Initialize: $weights(w^1) = 1/(M * N)$

for $t \leftarrow 1$ *to* $\mathcal{T}$ **do**

  Train sub-model θ_t with weights (w^t):

  $\mathcal{L}^t = \sum_{(u,i) \in \mathcal{O}} w^t_{u,i} \times \mathrm{L}^t_{u,i} / |\mathcal{O}|$

  Calculate $\mathcal{L}^t_{u,i}$ for each user-item pair;

  Calculate the ensemble weight of each user-item pair:

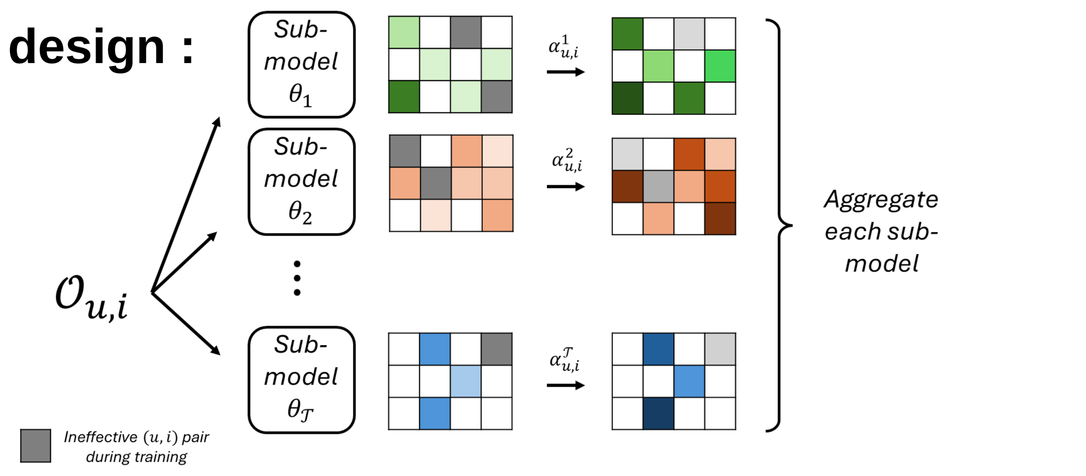
  $\alpha^t_{u,i} = \ln \frac{1}{\mathcal{L}^t_{u,i} + b}$

  Calculate weight of each user-item pair for the next iteration: $w^{t+1}_{u,i} = w^t_{u,i} \times e^{\alpha^t_{u,i} \mathcal{L}^t_{u,i}} / Z_t$;

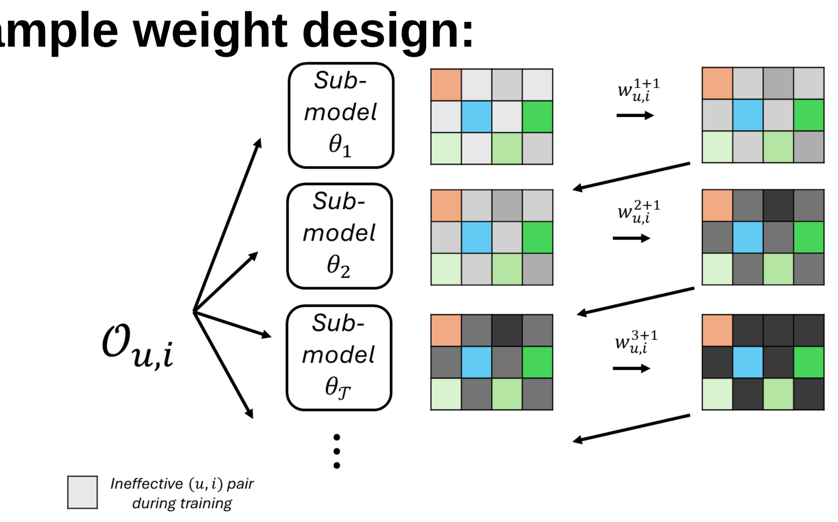
end for

Compute the final prediction: $\widehat{\mathcal{O}}_{u,i} = \sum_{t=1}^{\mathcal{T}} \alpha^t_{u,i} \times \widehat{\mathcal{O}}^t_{u,i}$.

- **$\alpha^t_{u,i}$ design :**



- **Sample weight design:**



Experiment Setup

- **Three Datasets:**

	#users	#items	density
KuaiRec	5,765	5,800	4.39%
Yelp	20,001	7,643	0.32%
CDs & Vinyl	12,023	8,050	0.32%

- **Evaluation Metrics:**
NDCG@K (user-oriented metrics), MDG@K (item-oriented metrics)
- **Baselines:**
MF, BPR, MultVAE (conventional algorithms without debiasing), LOCA, LFT, EnLFT (user-side debiasing SOTA), PC, BC Loss, Zero Sum (item-side debiasing SOTA)
- **Data and code:** https://github.com/JP-25/CFBoost

Proposed Model vs. State-of-the-art Models

- **Observations:**

(i) CFBoost mitigates both user-side and item-side biases simultaneously, outperforming bias-unaware models and SOTA user-side/item-side debiasing methods;
(ii) CFBoost has a trade-off between minority and privileged groups with suboptimal utility for the most privileged groups.

Table 1: Comparison across SOTA debiasing baselines and the proposed CFBoost on CDs & Vinyl (user mainstream bias).

	NDCG @20	Subgroups of user mainstream levels L	ML	M	MH	H
MF	.1292	.1095	.1171	.1330	.1388	.1474
BPR	.1055	.0917	.0986	.1112	.1113	.1149
MultVAE	.1382	.1175	.1244	.1446	.1477	.1570
LOCA	.1593	.1364	.1479	.1596	.1728	.1799
EnLFT	.1519	.1252	.1369	.1527	.1657	.1788
LFT	.1583	.1322	.1430	.1563	.1747	.1851
PC	.1021	.0884	.0965	.1064	.1078	.1116
BC Loss	.1360	.1135	.1275	.1429	.1437	.1521
Zero Sum	.1006	.0843	.0933	.1054	.1075	.1123
CFAdaBoost	**.1660**	**.1481**	.1532	**.1681**	**.1755**	**.1852**
CFBoost	.1644	.1472	**.1541**	.1658	.1698	.1851
Δ_{best} (%)	3.20	7.88	4.19	3.87	-1.74	0
Avg Δ_{best} (%)		2.84				

L: low, ML: med-low, M: medium, MH: med-high, H: high

Table 2: Comparison across SOTA debiasing baselines and the proposed CFBoost on CDs & Vinyl (user activeness bias).

	NDCG @20	Subgroups of user activeness levels L	ML	M	MH	H
MF	.1292	.1101	.1157	.1255	.1358	.1587
BPR	.1055	.0962	.1005	.1074	.1053	.1183
MultVAE	.1382	.1171	.1245	.1338	.1438	.1721
LOCA	.1593	.1331	.1424	.1535	.1675	.2003
EnLFT	.1519	.1259	.1304	.1452	.1582	.1995
LFT	.1583	.1318	.1396	.1529	.1644	**.2027**
PC	.1021	.0958	.1002	.1049	.1019	.1078
BC Loss	.1360	.1184	.1338	.1348	.1422	.1505
Zero Sum	.1006	.0880	.0929	.1028	.1021	.1169
CFAdaBoost	**.1660**	.1413	**.1506**	.1618	.1741	.2024
CFBoost	.1644	**.1416**	.1505	**.1637**	**1785**	.1876
Δ_{best} (%)	3.20	6.40	5.71	6.63	6.55	-7.45
Avg Δ_{best} (%)		3.57				

L: low, ML: med-low, M: medium, MH: med-high, H: high

Table 3: Comparison across SOTA debiasing baselines and the proposed CFBoost on CDs & Vinyl (item mainstream bias).

	MDG @20	Subgroups of item mainstream levels L	ML	M	MH	H
MF	.0266	.0206	.0172	.0205	.0276	.0473
BPR	.0088	.0032	.0018	.0047	.0093	.0252
MultVAE	.0234	.0096	.0091	.0158	.0272	.0554
LOCA	.0427	.0305	.0284	.0345	.0472	.0731
EnLFT	.0433	.0265	.0285	.0350	.0490	.0773
LFT	.0457	.0259	.0301	.0386	.0527	.0810
PC	.0143	.0293	.0070	.0063	.0081	.0209
BC Loss	.0453	.0478	.0381	.0399	.0446	.0561
Zero Sum	.0083	.0014	.0010	.0027	.0095	.0267
CFAdaBoost	.0371	.0306	.0249	.0290	.0381	.0629
CFBoost	**.0613**	**.0482**	**.0494**	**.0597**	**.0674**	**.0821**
Δ_{best} (%)	34.14	0.73	29.66	49.35	27.85	1.39
Avg Δ_{best} (%)		21.80				

L: low, ML: med-low, M: medium, MH: med-high, H: high

Table 4: Comparison across SOTA debiasing baselines and the proposed CFBoost on CDs & Vinyl (item popularity bias).

	MDG @20	Subgroups of item popularity levels L	ML	M	MH	H
MF	.0266	.0079	.0115	.0165	.0237	.0736
BPR	.0088	.0005	.0006	.0008	.0016	.0406
MultVAE	.0234	.0014	.0029	.0065	.0175	.0908
LOCA	.0427	.0143	.0233	.0281	.0418	.1063
EnLFT	.0433	.0116	.0202	.0273	.0438	.1135
LFT	.0457	.0109	.0206	.0319	.0501	**.1147**
PC	.0143	.0158	.0088	.0061	.0053	.0355
BC Loss	.0453	.0242	.0338	.0365	.0458	.0862
Zero Sum	.0083	.0001	.0002	.0003	.0006	.0409
CFAdaBoost	.0371	.0138	.0195	.0243	.0324	.0954
CFBoost	**.0613**	**.0458**	**.0554**	**.0590**	**.0588**	.0876
Δ_{best} (%)	34.14	89.57	63.98	61.62	17.36	-23.65
Avg Δ_{best} (%)		41.78				

L: low, ML: med-low, M: medium, MH: med-high, H: high

Conclusion and Future Work

- **Conclusions:**

(i) Propose the CFBoost, to address a broad spectrum of model biases.

(ii) Theoretical analysis shows that the model can achieve an exponential dropping upper bound on the training loss for any user-item pairs with increasing boosting iterations.

(iii) Extensive experiments demonstrate the superior performance of our proposed method for both privileged and minority groups compared to SOTA alternatives.

- **Future Work:**

Exploring CFBoost's debiasing efficacy across a broader spectrum of model biases beyond the four studied in this work.