RecSys 2025
Prague

Leave No One Behind: Fairness-Aware Cross-Domain Recommender Systems for Non-Overlapping Users

Weixin Chen*
Hong Kong Baptist University, China
cswxchen@comp.hkbu.edu.hk

Yuhan Zhao*
Hong Kong Baptist University, China
csyhzhao@comp.hkbu.edu.hk

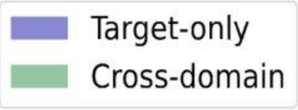
Li Chen
Hong Kong Baptist University, China
lichen@comp.hkbu.edu.hk

Weike Pan
Shenzhen University, China
panweike@szu.edu.cn

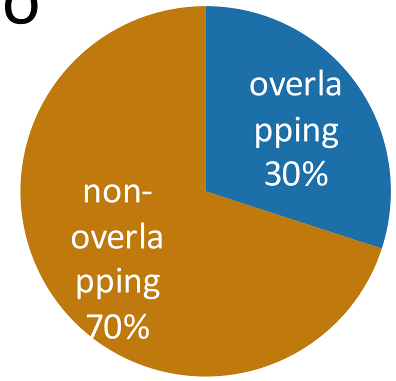
*Equal contributions.

Motivation

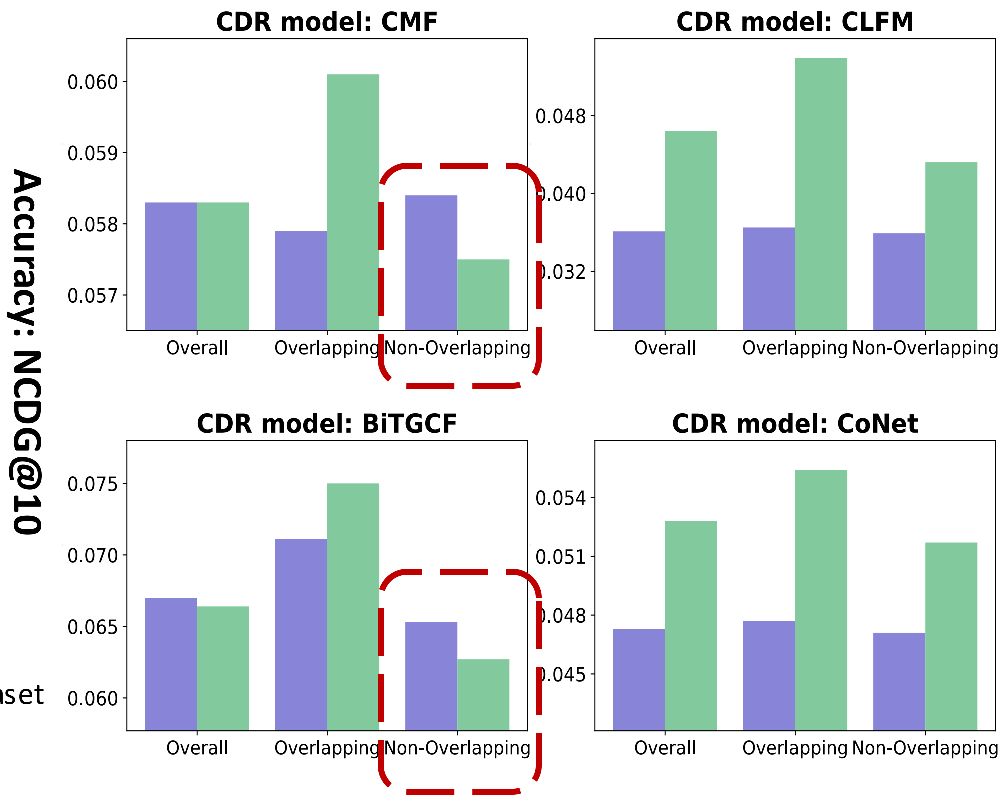
- **Cross-domain recommendation (CDR):** Leveraging **overlapping users** as a knowledge bridge to promote accuracy in target domain



- **CDR overlapping bias: non-overlapping users** see less improvement or even performance degradation compared to **overlapping users.**



Epinions dataset



- **Unfairness** in CDR!

- We validate CDR overlapping bias and provide theoretical analyses in our paper.

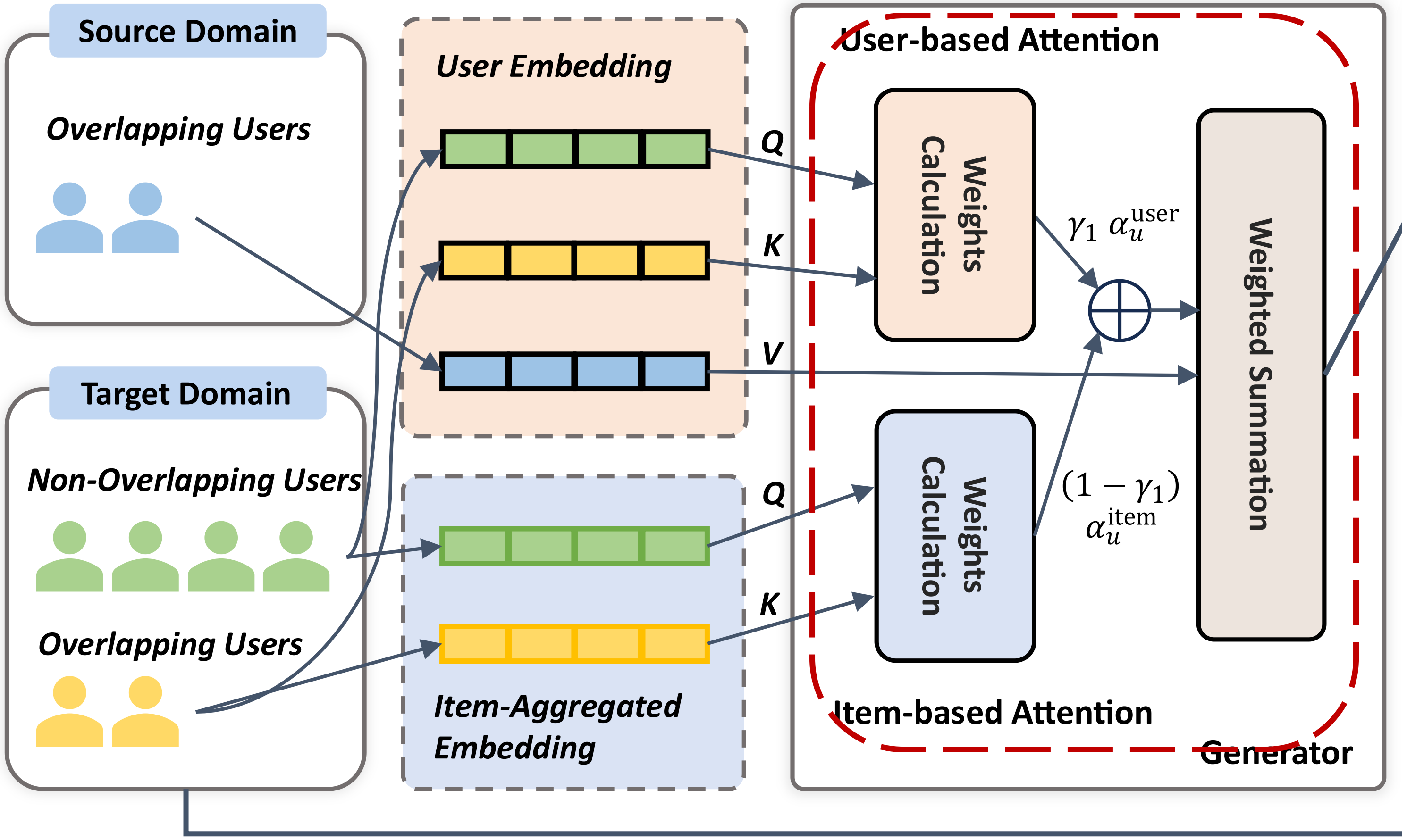
The recommendation quality for non-overlapping users **DEGRADES** after cross-domain learning on Epinions dataset.

Motivation

- **An intuitive solution:**
  - Let all users enjoy cross-domain learning as **overlapping users**!
  - **Generate source embeddings** for non-overlapping users in the target domain.

- **But it is technically challenging...**
  - (*C1*) How to generate virtual users for non-overlapping target users with **no existing data** in the source domain?
  - (*C2*) How to ensure that the virtual users generated remain representative of the **real source-domain data distribution**?

Our Method: VUG (Virtual User Generation)



- (*C1*) How to generate virtual users for non-overlapping target users with **no existing data** in the source domain?

- Leveraging similar overlapping users:

$$\{u_1, u_2, \ldots, u_N\} = \underset{u \in \mathcal{U}^o}{\arg\text{top}\,N}\, \text{Sim}\left(\mathbf{e}_{non}^{T}, \mathbf{e}_u^{T}\right)$$

$$\mathbf{e}_{\text{non}}^{S'} = \text{Aggr}\left(\mathbf{e}_u^{S}\right), \quad u \in \{u_1, u_2, \ldots, u_N\},$$

For *C1*: Generate source embeddings based on similar overlapping users and items

Our Method: VUG (Virtual User Generation)

For *C2*: Introduce a limiter with supervision loss and constraint loss to ensure generation fidelity

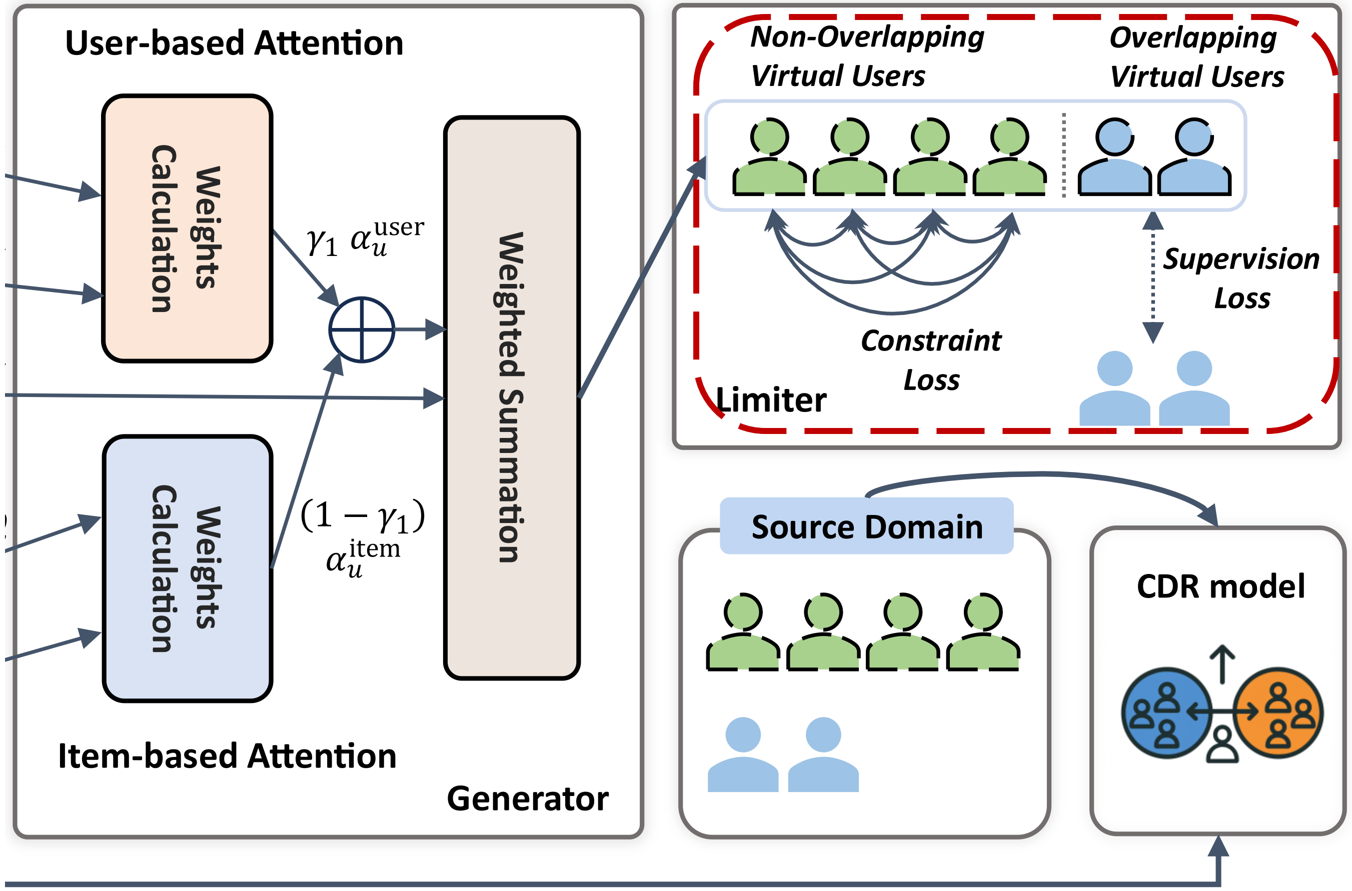
- (*C2*) How to ensure that the virtual users generated remain representative of the **real source-domain data distribution**?

- Supervision loss over overlapping users:

$$\mathcal{L}_{\text{super}} = \frac{1}{|\mathcal{U}^o|} \sum_{o \in \mathcal{U}^o} \left\| \mathbf{e}_o^{S'} - \mathbf{e}_o^{S} \right\|^2$$

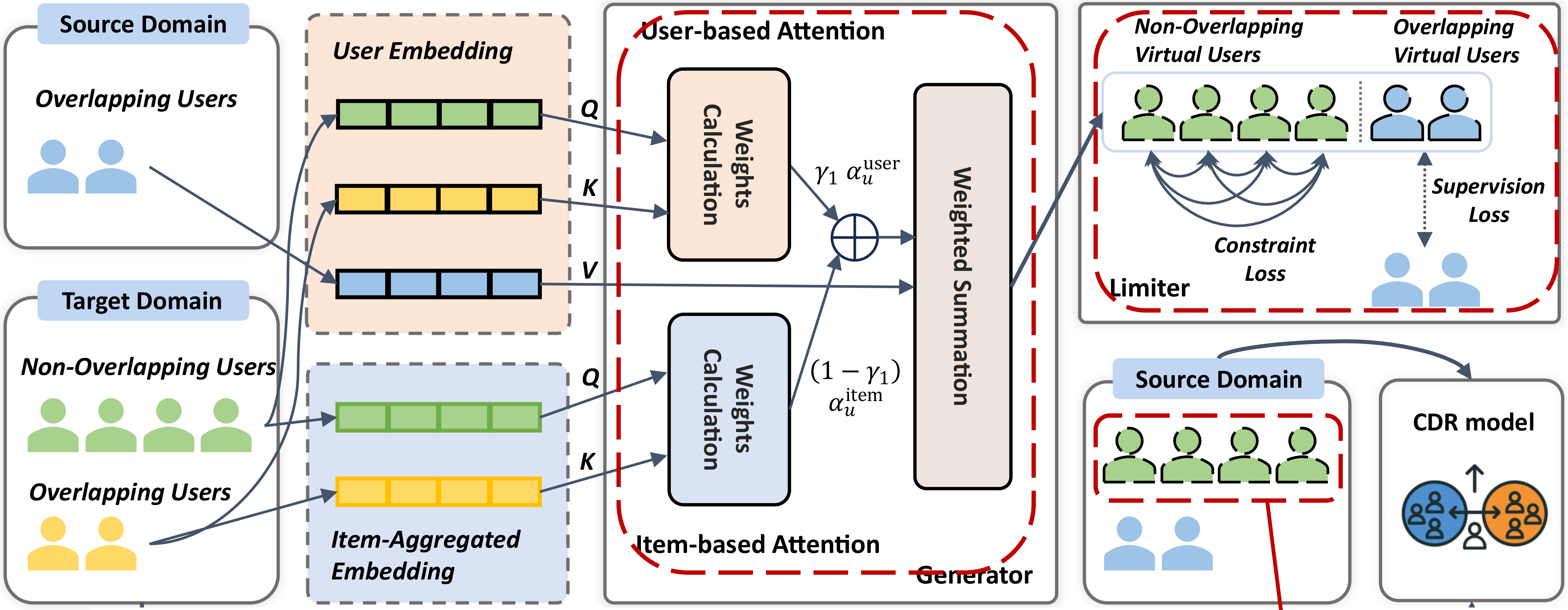
- Constraint loss among non-overlapping users:

$$\mathcal{L}_{\text{constrain}} = \log \mathbb{E}\left[e^{-2\left\| \mathbf{e}_u^{S'} - \mathbf{e}_{u'}^{S'} \right\|^2} \right]$$



Our Method: VUG (Virtual User Generation)

For *C2*: Introduce a limiter with supervision loss and constraint loss to ensure generation fidelity



For *C1*: Generate source embeddings based on similar overlapping users and items

Generated source embeddings for the target-domain non-overlapping users

Experimental Setting

- On three real-world datasets

Dataset	Domain	#Users	#Items	#Inter.	Overlap
Amazon	Book	65,725	70,071	2,021,443	2.55%
	Movie	8,663	7,790	229,257	19.32%
Douban	Book	18,086	33,067	809,248	5.94%
	Movie	3,372	9,342	311,797	31.88%
Epinions	Elec	10,124	13,018	34,859	12.51%
	Game	4,247	4,094	16,471	29.83%

- As our VUG is **model-agnostic**, we examine its performance by integrating it into widely-used CDR baselines.

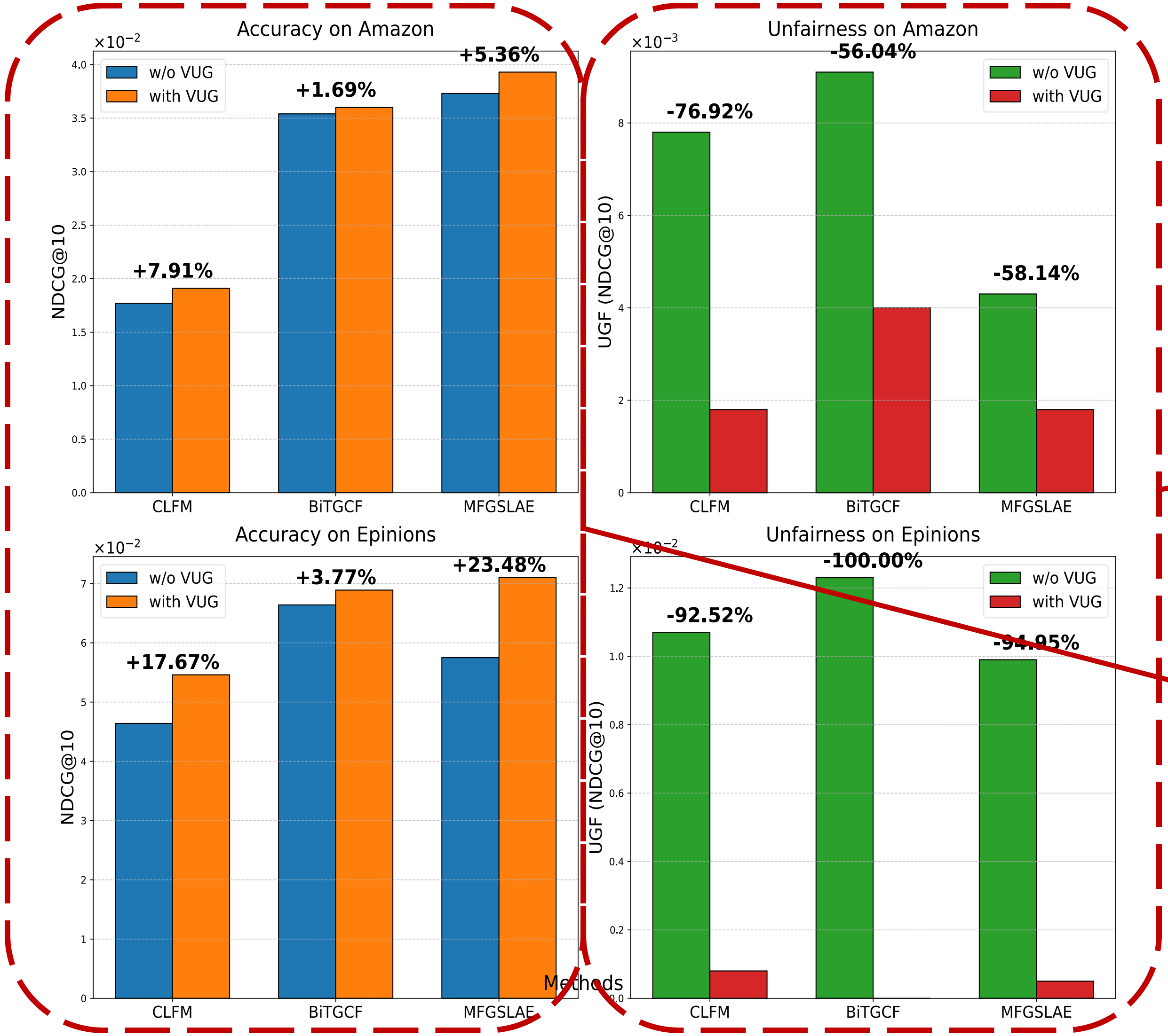
- Accuracy metrics: NCDG and HR

- Fairness metrics

$$\mathrm{UGF} = \left| \frac{1}{|\mathcal{U}^o|} \sum_{u \in \mathcal{U}^o} \mathcal{M}(L_u) - \frac{1}{|\mathcal{U}^T \setminus \mathcal{U}^o|} \sum_{u \in \mathcal{U}^T \setminus \mathcal{U}^o} \mathcal{M}(L_u) \right|$$

UGF measures **absolute accuracy difference** (*e.g.*, NCDG@10) between overlapping and non-overlapping users.

Major Results



UGF measures **absolute accuracy difference** (*e.g.*, NCDG@10) between overlapping and non-overlapping users. **Lower UGF, better fairness performance.**

Our method VUG largely improves the **fairness** performance.

Our method VUG also slightly improves the **overall accuracy** performance.

Ablation Study

Method	Accuracy (*larger* is better)			
	HR@10	HR@20	NDCG@10	NDCG@20
w/o $\mathcal{L}_{constrain}$	0.1302	0.1974	0.0655	0.0821
w/o $\mathcal{L}_{super}$	0.1355	0.1985	0.0707	0.0864
w/o α_u^{user}	0.1298	0.1974	0.0662	0.0828
w/o α_u^{item}	0.1287	0.1963	0.0658	0.0824
VUG	**0.1363**	**0.2001**	**0.0710**	**0.0865**

Method	UGF (*smaller* is better)			
	HR@10	HR@20	NDCG@10	NDCG@20
w/o $\mathcal{L}_{constrain}$	0.0010	0.0168	0.0005	0.0033
w/o $\mathcal{L}_{super}$	0.0281	0.0152	0.0129	0.0096
w/o α_u^{user}	0.0020	0.0095	0.0015	0.0034
w/o α_u^{item}	0.0041	0.0112	**0.0005**	0.0035
VUG	**0.0003**	**0.0015**	**0.0005**	**0.0014**

We examine the contributions of each component of VUG, in terms of accuracy and fairness.

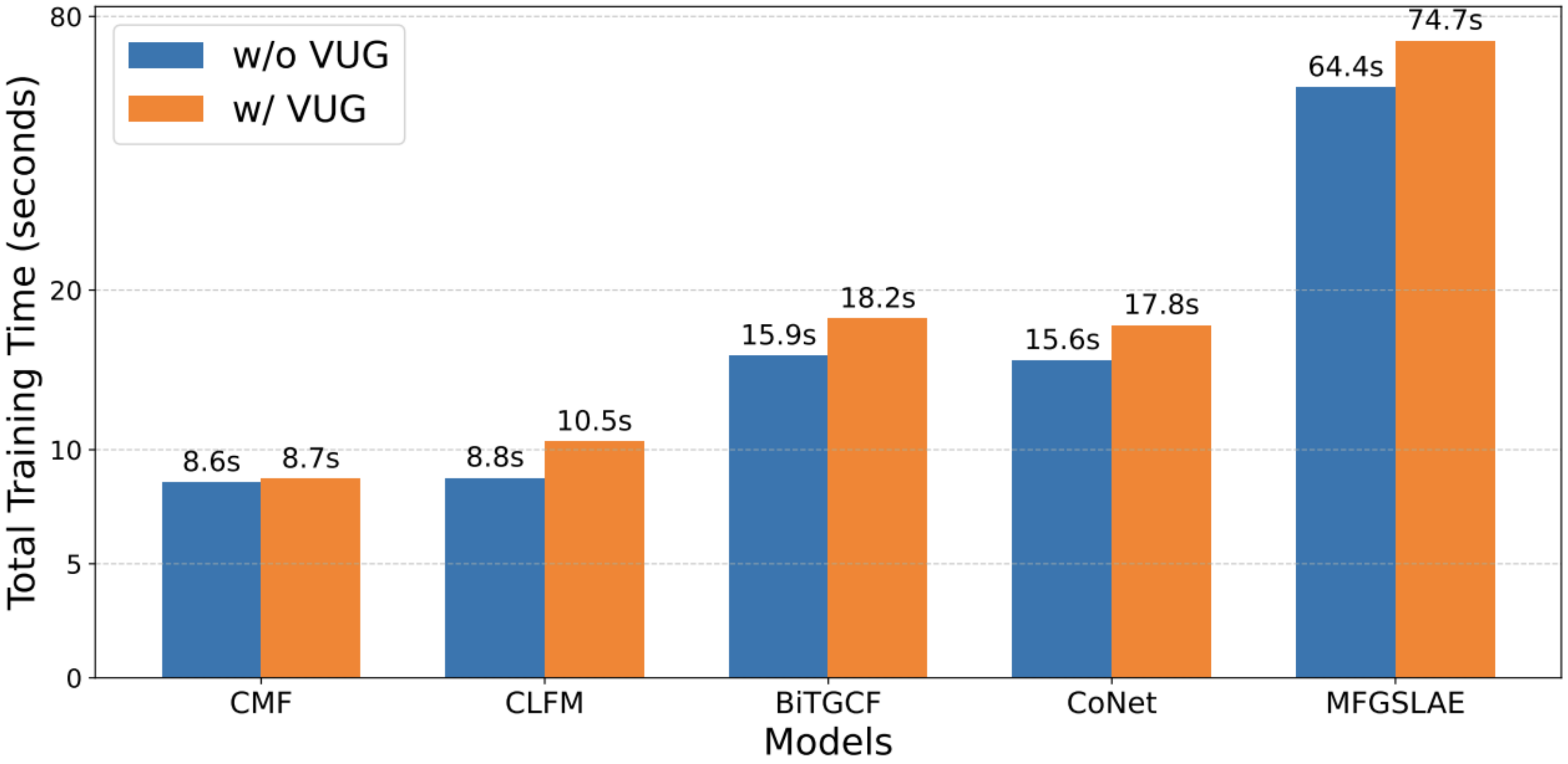
Accuracy: HR and NDCG

Unfairness: UGF

All components positively contribute to VUG's accuracy and fairness performance.

Efficiency Analysis

Some may concern that integrating VUG makes cross-domain recommendation models too complex and **inefficient**...



Conclusion and Future Work

- **Conclusion:**
  - We address a critical unfairness issue in cross-domain recommendations.
  - We propose a **virtual user generation (VUG)** framework that generates source embeddings for non-overlapping users in the target domain.
- **Future work:**
  - To explore unfair phenomenon exists in other CDR methods, such as review-based approaches.
  - To explore the application and deployment of VUG within real-world industrial scenarios.

RecSys 2025
Prague

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

深圳大学
SHENZHEN UNIVERSITY

Thank you!

Our group

Code