

# Leave No User Behind: Towards Improving the Utility of Recommender Systems for Non-mainstream Users

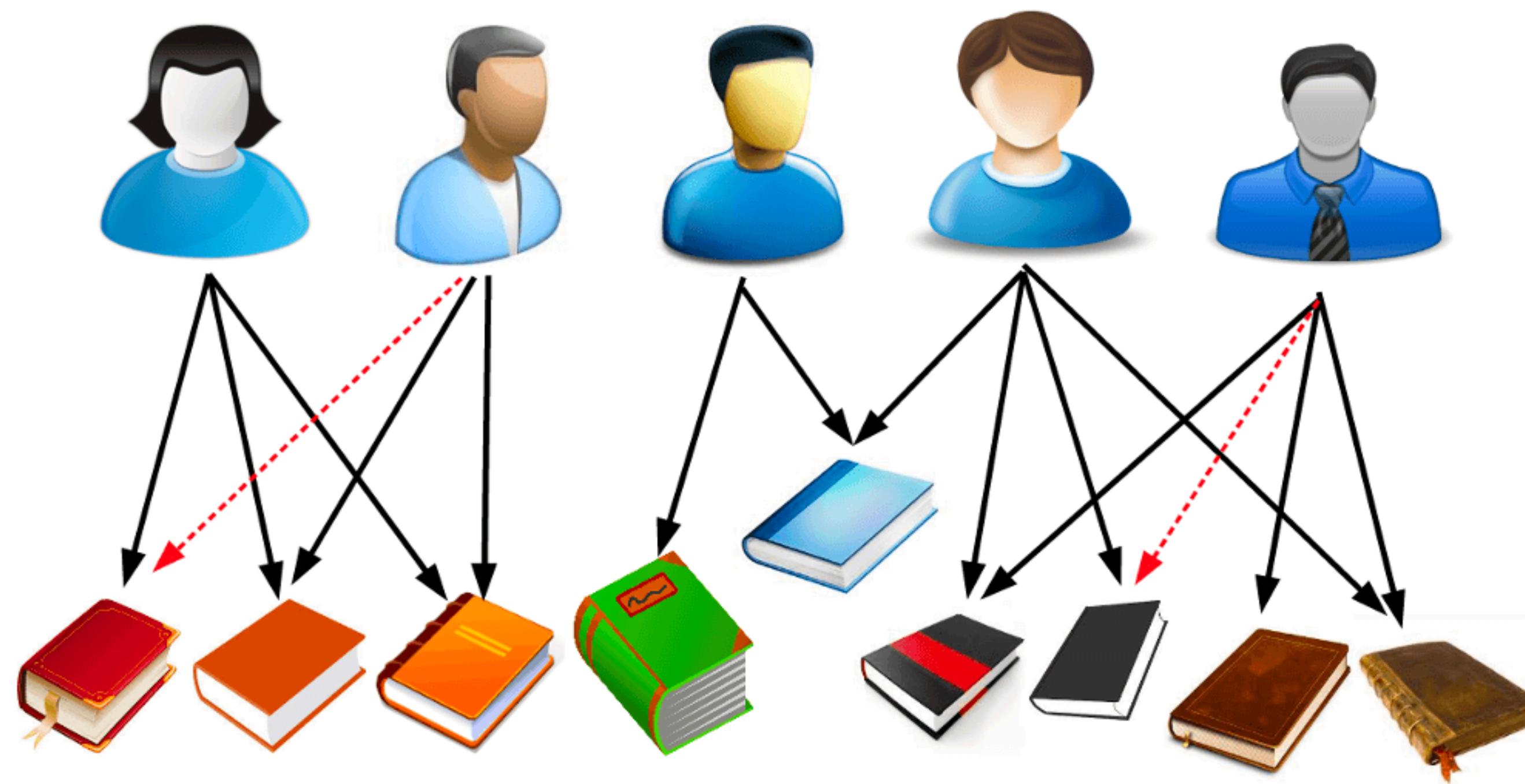
Roger Zhe Li (Presenter), Julián Urbano, Alan Hanjalic

Delft University of Technology, the Netherlands

Email: [z.li-9@tudelft.nl](mailto:z.li-9@tudelft.nl)

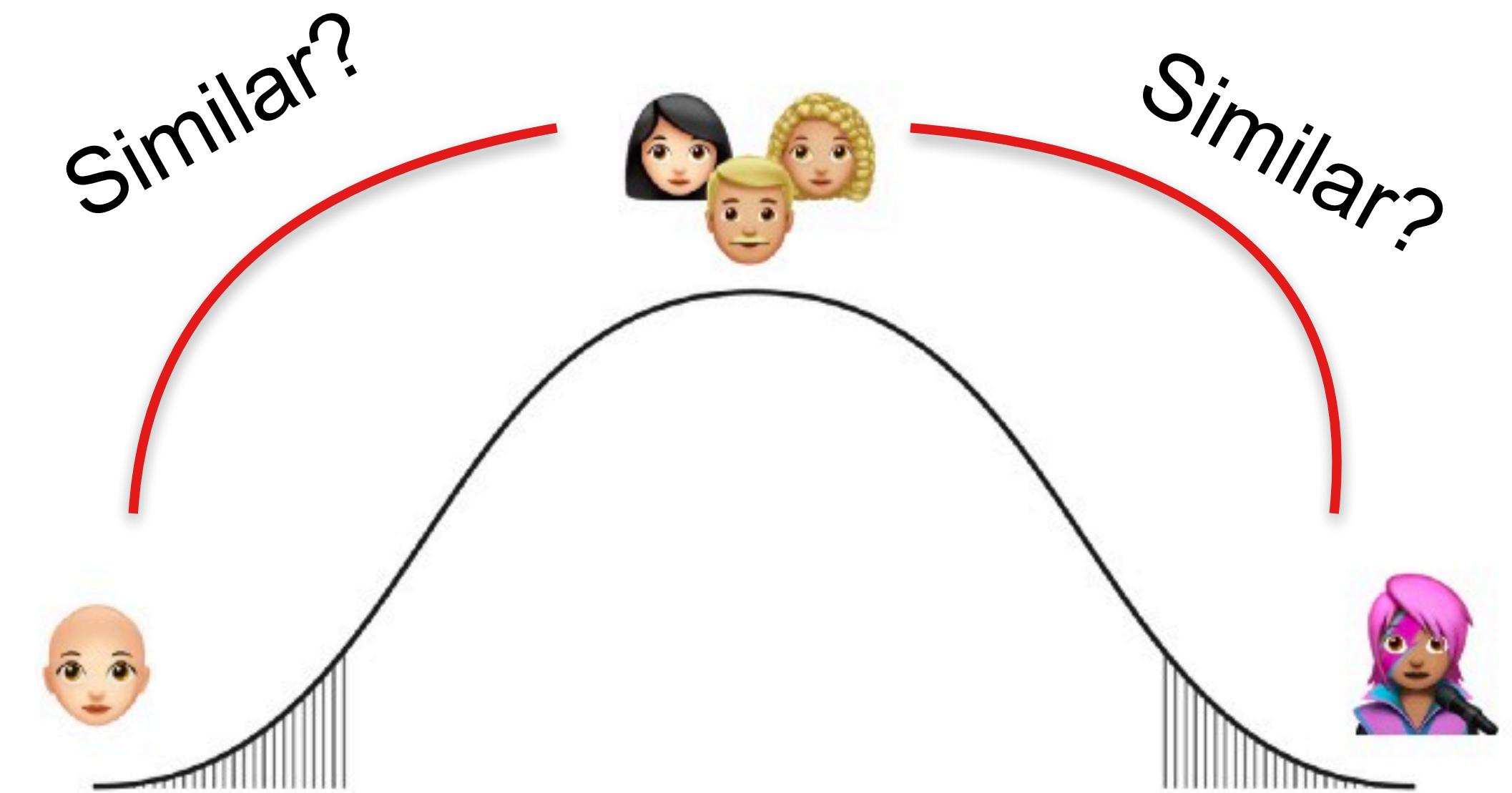


# Motivation: Collaborative Filtering



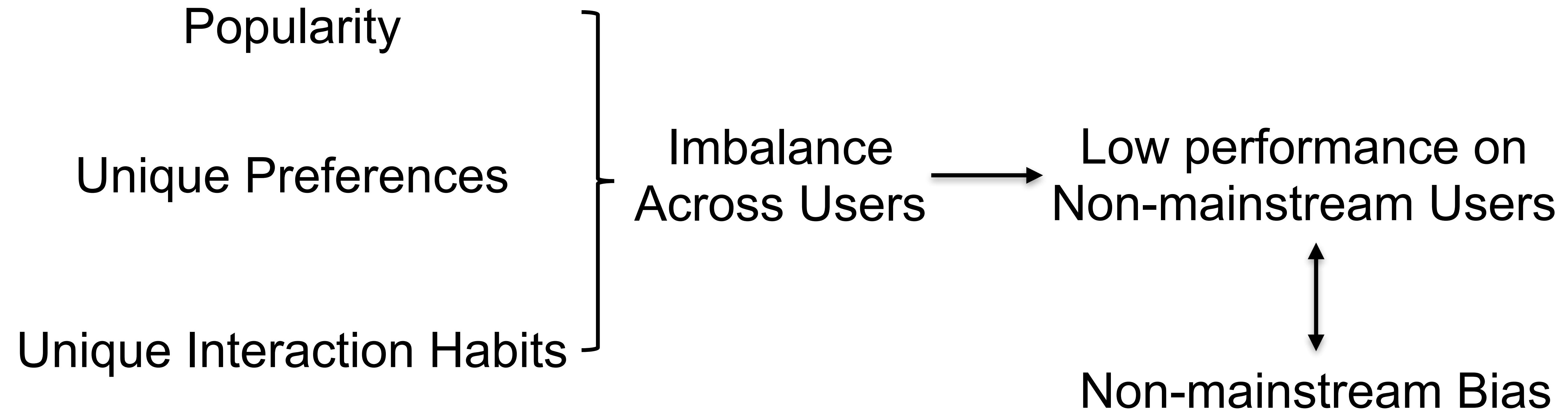
CF is similarity-based.

# Similarity, Are We Going Too Far?



<https://uxdesign.cc/the-fundamentals-of-engaging-with-extreme-users-45e0033e6b2>

# Biases in CF-based Recommendation



# Strategies

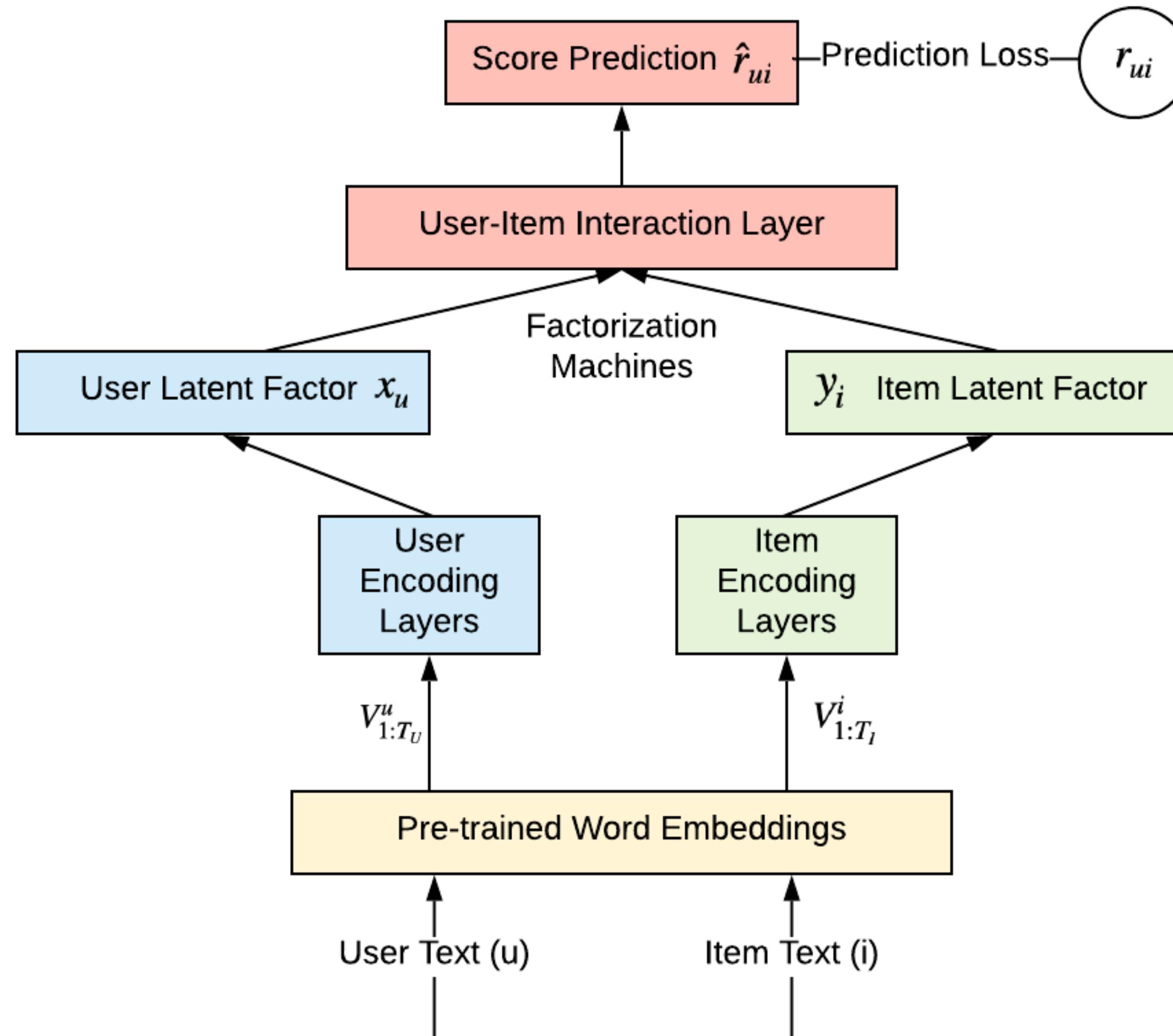
- Using content features to augment the information of non-mainstream users;
- Introducing an adversarial power to prevent pure similarity-oriented learning, and thus preserve the unique user and item properties.

# Problem

- Goal: alleviating the non-mainstream bias.
- Given: explicit ratings; content-based side information ([review texts](#)).
- Target: A personalized recommendation list to each user, achieving a **better balance** across all users in recommendation.

# How to utilize the content-based side information?

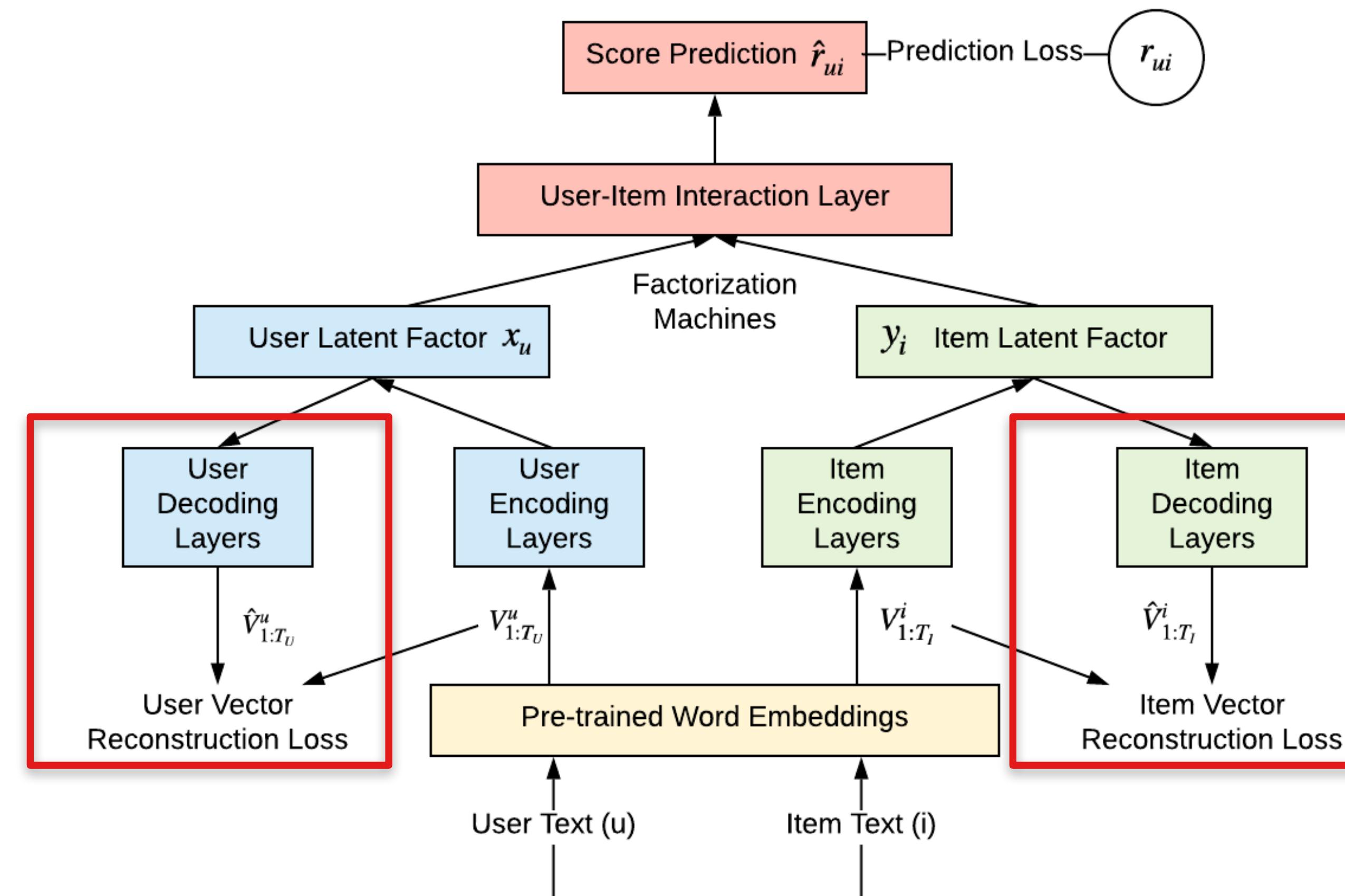
# DeepCoNN: Preliminary Model



Zheng, Lei, Vahid Noroozi, and Philip S. Yu. "Joint deep modeling of users and items using reviews for recommendation." *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017

# How to provide the adversarial power?

# NAECF: Debiasing for Non-Mainstream Users



The architecture of *Neural AutoEncoder Collaborative Filtering*

# Model Learning: Overall Loss

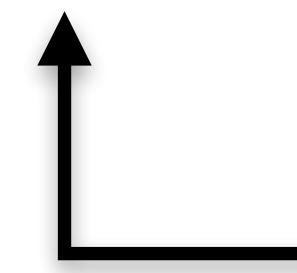
The overall loss function is the linear combination generated from CF (rating prediction) and AE (text reconstruction).

$$L = L_R + w (L_U + L_I)$$



Trade-off between CF and AE

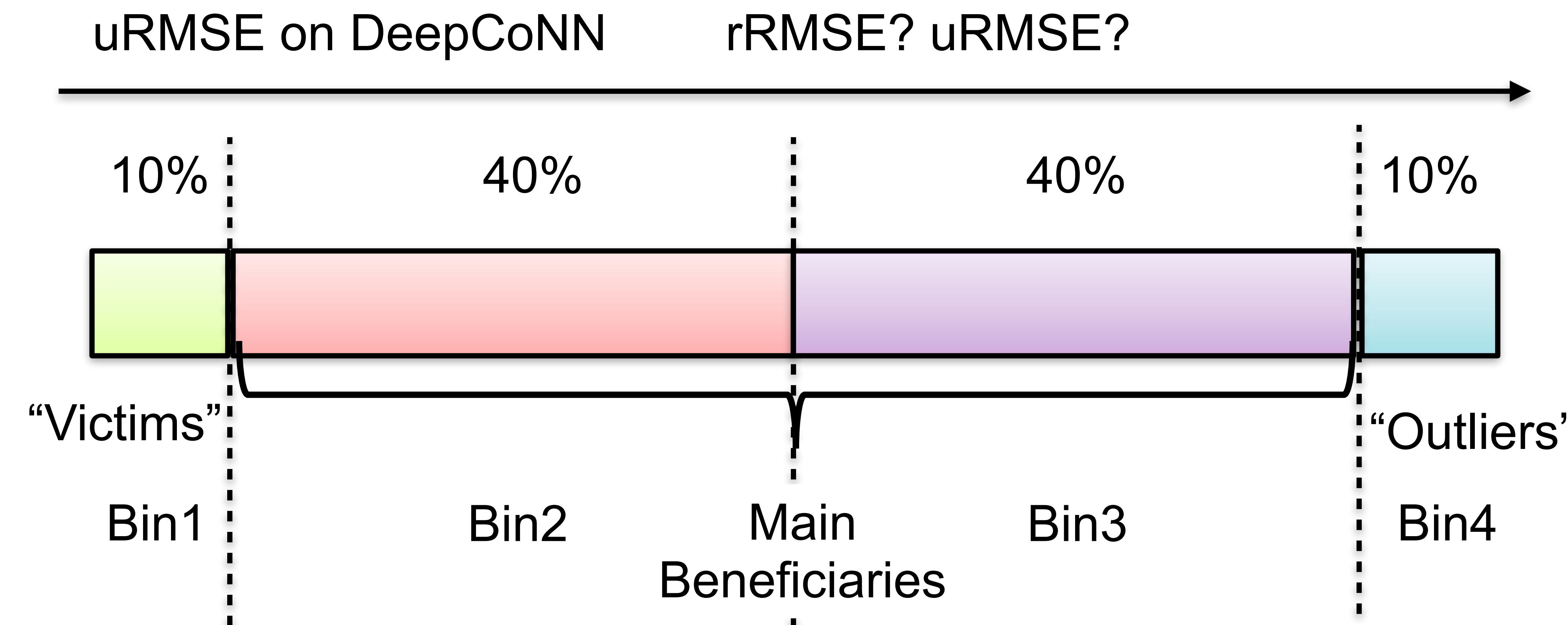
$$\{0, 0.1, 0.2, 0.5, 1, 2, 5, 10\}$$



Equivalent to DeepCoNN

# How to evaluate the balance across users?

# Individual User Investigation: AutoEncoders

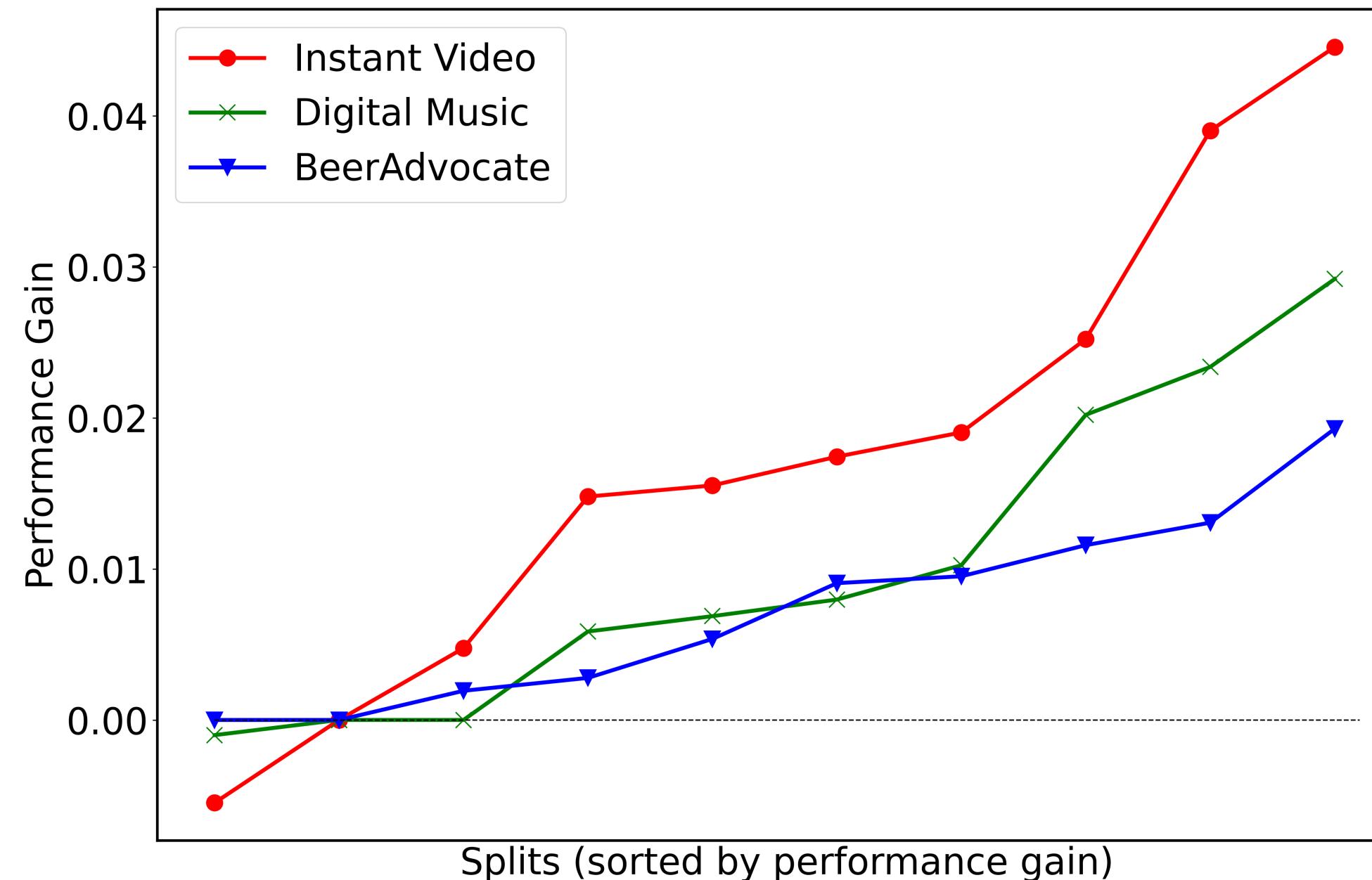


# Experiments: Datasets & Baselines

Dataset	#users	#items	#ratings	Sparsity	#words
Instant videos	5,130	1,685	37,126	99.57	19M
Digital music	5,541	3,568	64,706	99.67	73M
BeerAdvocate	3,703	37,580	393,035	99.72	198M

- Train: Validation: Test = 8:1:1
- 10 different splits per dataset
- **Core Evaluation Metric:** rRMSE, uRMSE
- Each user has at least 5 ratings
- Baselines: MF and DeepCoNN

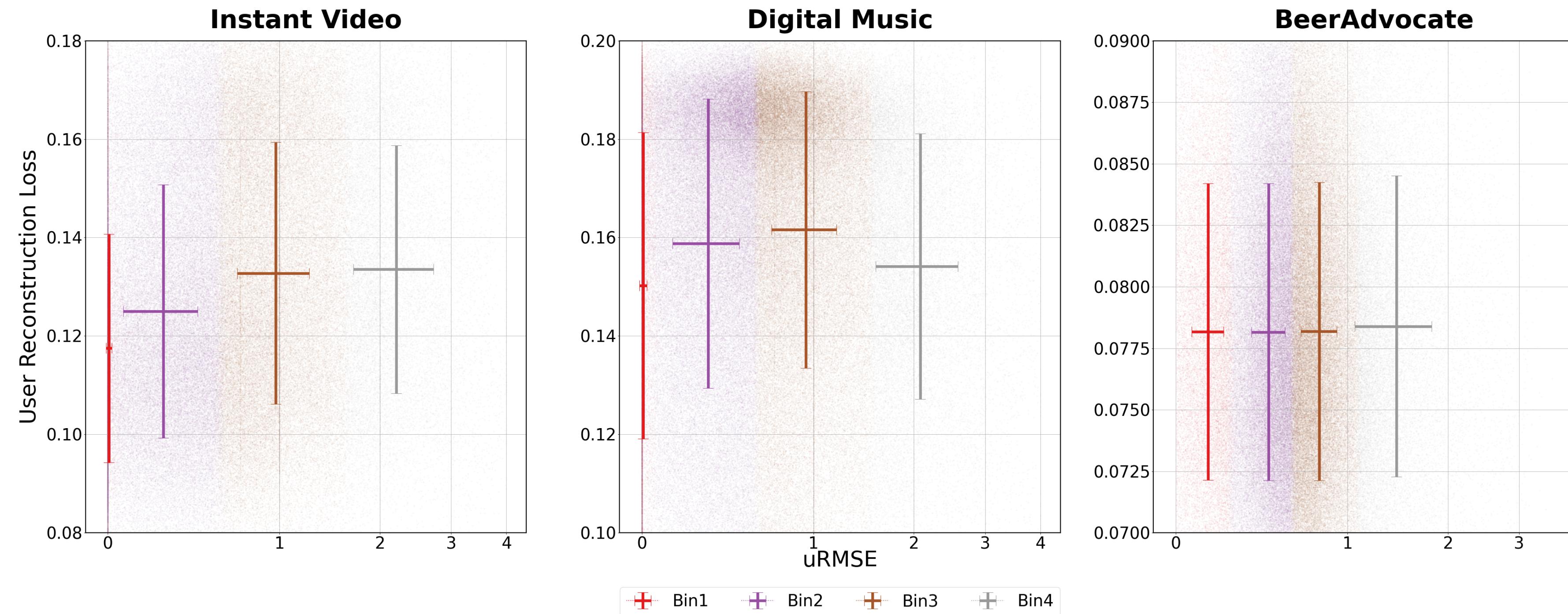
# Individual User Investigation: AutoEncoders



- NAECF improves the utilities of non-mainstream users WITHOUT significantly hurting mainstream users.

Dataset	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$	$\Delta$
Instant Video	-0.0035	0.0256*	0.0267*	-0.0308*	0.0175*
Digital Music	0.0036	0.0184*	0.0106*	-0.0167*	0.0103*
BeerAdvocate	0.0119	0.0117*	0.0063*	-0.0115*	0.0073*

# Individual User Investigation: Mechanisms



- The vectors of Mainstream users are easier to reconstruct.

Dataset	$\Delta_1$	$\Delta_2$	$\Delta_3$	$\Delta_4$	$\Delta$
Instant Video	-0.0035	0.0256*	0.0267*	-0.0308*	0.0175*
Digital Music	0.0036	0.0184*	0.0106*	-0.0167*	0.0103*
BeerAdvocate	0.0119	0.0117*	0.0063*	-0.0115*	0.0073*

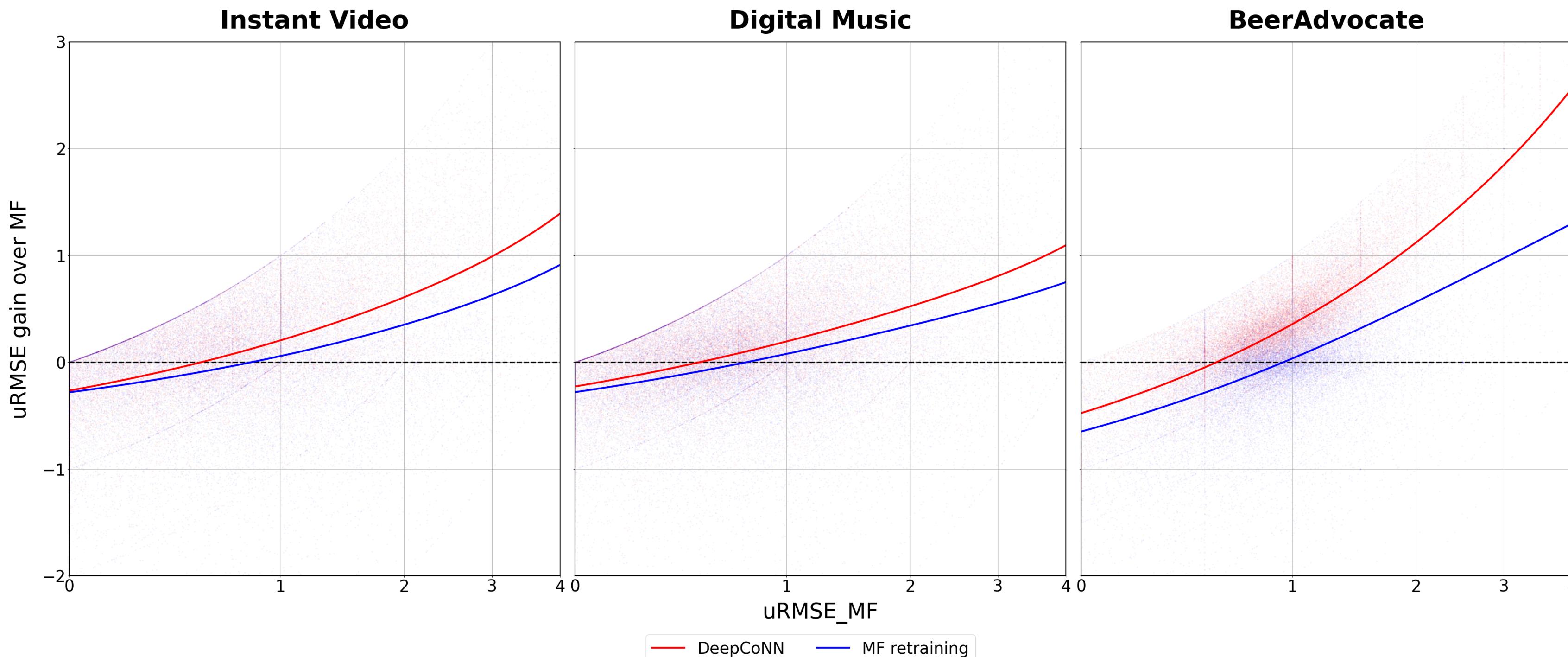
# Experiments: Overall Results

Table 2:  $rRMSE$  over 10 data splits for all recommendation methods in all three datasets ( $mean \pm std.dev.$ ). Weighted NAECFs are represented by the weights of the text rebuilding loss  $w$ . Best results per dataset are in bold. Results statistically different from the best ( $t$ -test,  $p < 0.05$ ) are marked with \*.

Dataset	MF	DeepCoNN	NAECF						
			0.1	0.2	0.5	1.0	2.0	5.0	10.0
Instant Video	$1.1600 \pm .0264^*$	$0.9744 \pm .0145$	<b><math>0.9732 \pm .0149</math></b>	$0.9749 \pm .0122$	$0.9754 \pm .0159$	$0.9757 \pm .0169$	$0.9798 \pm .0212$	$0.9896 \pm .0221^*$	$0.9967 \pm .0221^*$
Digital Music	$1.0466 \pm .0097^*$	<b><math>0.9078 \pm .0138</math></b>	$0.9083 \pm .0115$	$0.9106 \pm .0128$	$0.9097 \pm .0114$	$0.9104 \pm .0128$	$0.9118 \pm .0134$	$0.9167 \pm .0108$	$0.9219 \pm .0146^*$
BeerAdvocate	$1.0442 \pm .0048^*$	$0.6722 \pm .0090$	$0.6707 \pm .0064$	<b><math>0.6692 \pm .0035</math></b>	$0.6746 \pm .0059^*$	$0.6756 \pm .0082^*$	$0.6785 \pm .0098^*$	$0.6899 \pm .0137^*$	$0.7068 \pm .0278^*$

- Review features help achieve a significant improvement over pure MF.
- NAECF with a mild CF-AE trade-off does NOT bring a significant rRMSE loss.

# Individual User Investigation: Reviews



- Review features help achieve a significant improvement over pure MF on user balance.

# Conclusions

- Collaborative filtering methods suffer a non-mainstream bias across different users;
- With the adversarial power of AutoEncoders, NAECF achieves a better user balance on top of DeepCoNN.
- Reviews can help promote the recommendation utility for non-mainstream users;

Code & Data: <https://github.com/roger-zhe-li/wsdm21-mainstream>

Special thanks to SIGIR for providing a travel grant for the first author.