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

Roger Zhe Li, Julián Urbano, Alan Hanjalic

z.li-9@tudelft.nl, Multimedia Computing Group, TU Delft

Introduction

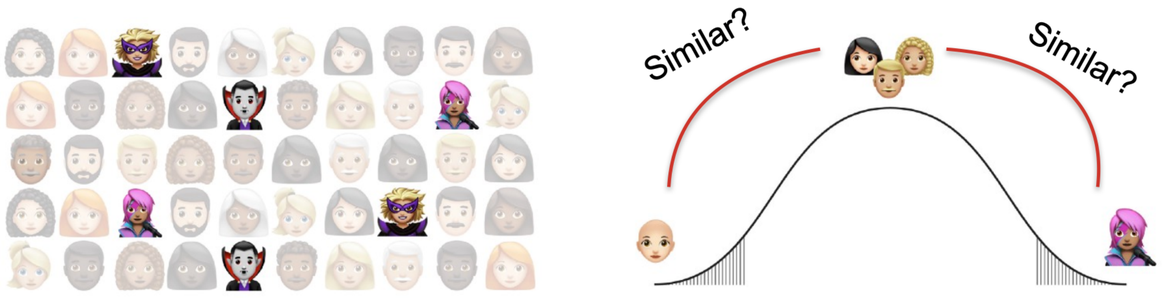

Figure 1: Non-mainstream users are under-represented in CF.

How to help non-mainstream users?

- Provide more information sources (review texts);
- Counteract the similarity-oriented learning process with AutoEncoders.

NAECF: Debiasing for Non-mainstream Users

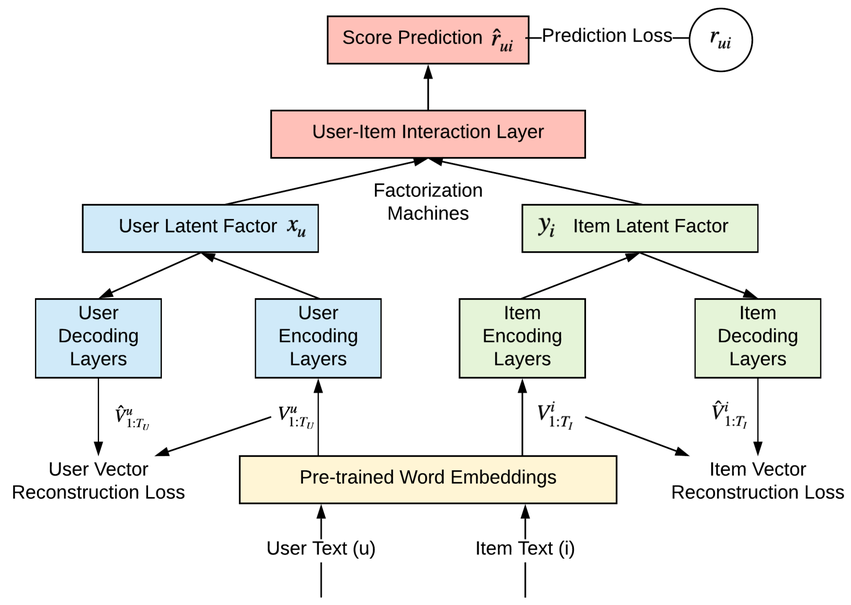

Figure 2: Overall architecture of NAECF.

Overall Loss:

$$L = L_R + w\left(L_U + L_I\right), \tag{1}$$

L_R: MSE prediction loss
L_U: Cos-similarity loss for user vector reconstruction
L_I: Cos-similarity loss for item vector reconstruction

Experimental Setups

Table 1: Statistics of the datasets.

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

- Experiment on 10 different splits for each dataset;
- Train:Validation:Test=8:1:1;
- w={0, 0.1, 0.2, 0.5, 1, 2, 5, 10};
- Evaluation Metrics: rRMSE (overall), uRMSE (individual);
- Users grouped as per their uRMSE scores on w=0 for validation (top 10%, 10%-50%, 50%-90%, >90%);
- Validation standard:

$$\Delta = 0.1\Delta_1 + 0.4\Delta_2 + 0.4\Delta_3 + 0.1\Delta_4. \tag{2}$$

Results

1. Effect of AutoEncoders

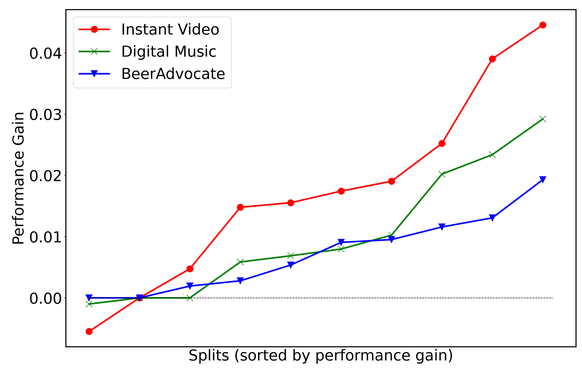

Figure 3: Test-set performance gain on each of 10 splits.

2. Mechanisms: Why NAECF Works?

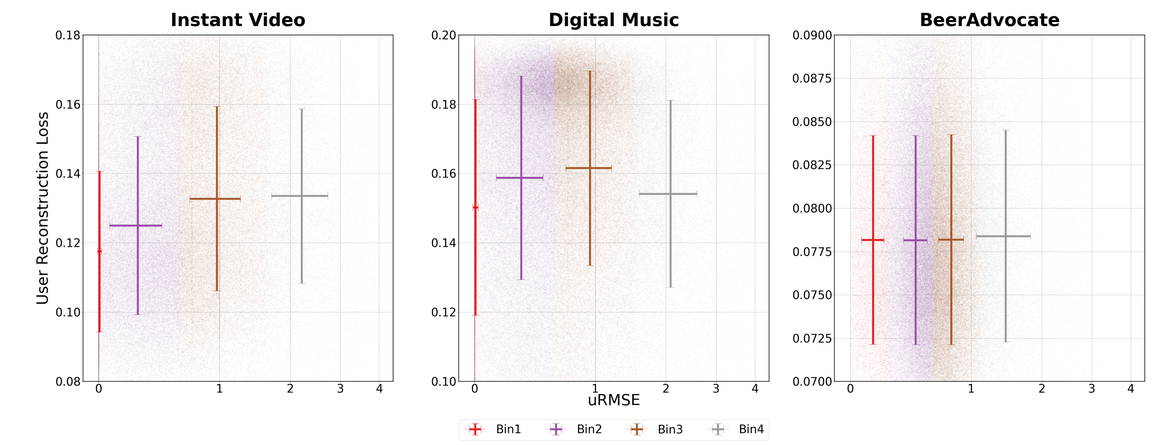

Figure 4: NAECF $loss_U$ by test-set $uRMSE$, with all evaluated weights w.

3. Power of Review Texts

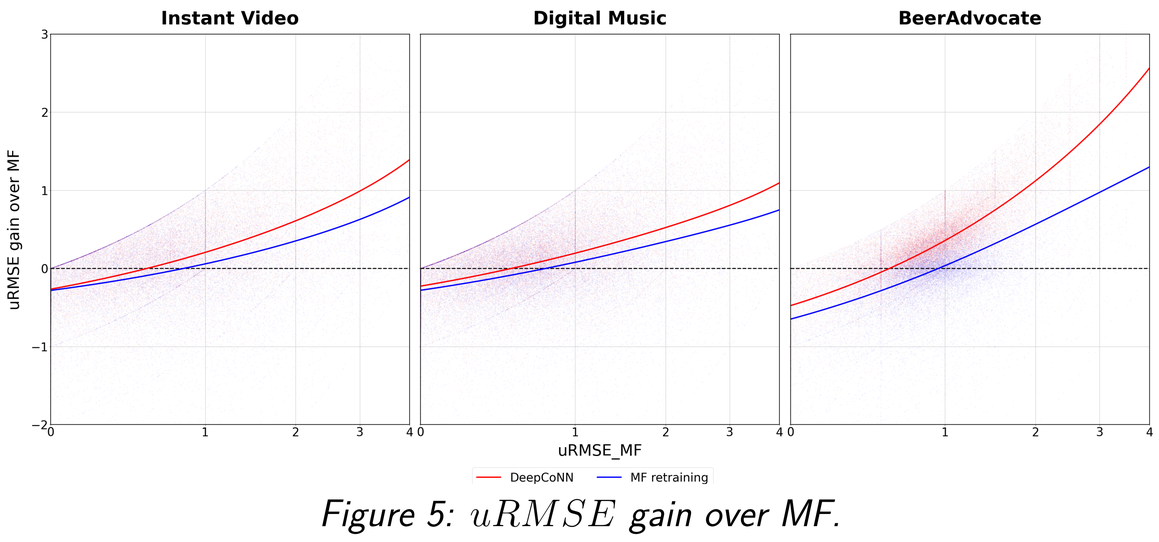

Figure 5: $uRMSE$ gain over MF.

Conclusions

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

Poster Available at

WSDM'21