Post-Training Embedding Enhancement for Long-Tail Recommendation (Supplementary Document)

Geon Lee
KAIST
Seoul, Republic of Korea
geonlee0325@kaist.ac.kr

Kyungho Kim
KAIST
Seoul, Republic of Korea
kkyungho@kaist.ac.kr

Kijung Shin
KAIST
Seoul, Republic of Korea
kijungs@kaist.ac.kr

ABSTRACT

In this supplementary document, we provide additional details and experimental results to support the main paper. The code and datasets are available at https://github.com/geon0325/EDGE.

ACM Reference Format:

Geon Lee, Kyungho Kim, and Kijung Shin. 2024. Post-Training Embedding Enhancement for Long-Tail Recommendation (Supplementary Document). In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/XXXXXX.XXXXXX

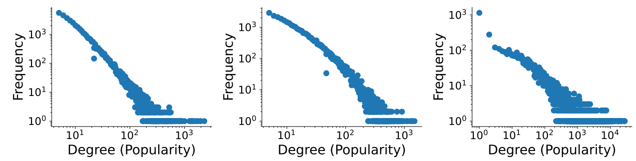
A ADDITIONAL ANALYSES & OBSERVATIONS

In this section, we present further observations that supplement those in the main paper.

Long-Tail Distributions. As shown in Figure 1, real-world datasets exhibit long-tail distributions in item popularity. A small fraction of items (head items) attracts most of the attention, while the majority (tail items) receive significantly less [5]. In these distributions, 20% of the items account for 80% of the interactions.

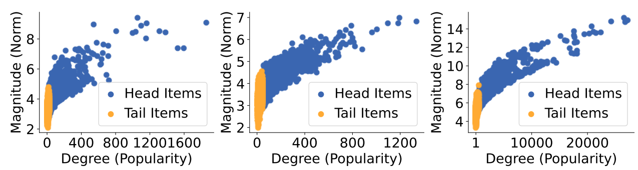
Magnitudes of Tail Item Embeddings. In Figure 2, we observe that tail items exhibit a weaker correlation between embedding magnitudes and popularity compared to head items. Here, the embeddings are obtained by LightGCN [2]. In Table 1, we examine the correlation between the magnitude of embeddings and item popularity for BPRMF [4] and LightGCN [2]. The results indicate that these models learn embeddings where the magnitude is more strongly correlated with the popularity of head items compared to tail items. This suggests that the magnitudes of embeddings for tail items are less effective at reflecting their popularity than those for head items. We conjecture that this discrepancy primarily stems from the limited data available for tail items, which challenges recommender systems to capture their popularity in the embedding magnitudes during training.

Directions of Tail Item Embeddings. In Figure 3, we examine the density of the distribution of the variance of the cosine similarity between each item embedding and user embeddings, i.e., $\mathrm{Var}(\cos(\theta_{1i}), \cdots, \cos(\theta_{|\mathcal{U}|i}))$. Here, the embeddings are obtained by LightGCN [2]. Intuitively, a higher variance indicates that the item embedding has greater discriminative power in identifying users who are close to the item and those who are not. However, we observe that tail items tend to have smaller variances in the angles with user embeddings, suggesting that tail items are less effective at capturing user preferences. In Table 2, we observe that both BPRMF [4] and LightGCN [2] exhibit a higher average variance for head items compared to tail items. We conjecture that this issue is also primarily rooted in the data scarcity associated with tail items.



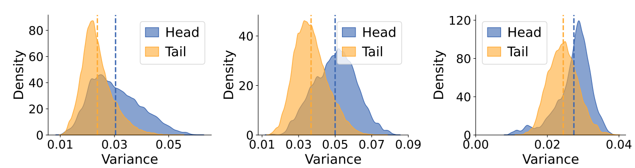
(a) Gowalla (b) Yelp (c) MovieLens

Figure 1: Long-tail distribution in real-world datasets. A few items (head) attract most of the attention, while the majority (tail) receive much less.



(a) Gowalla (b) Yelp (c) MovieLens

Figure 2: The magnitudes of embeddings for tail items are less correlated with their popularity than those for head items, indicating that the popularity of tail items is less reflected in the magnitudes of the embeddings.

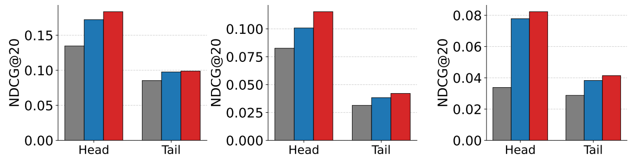


(a) Gowalla (b) Yelp (c) MovieLens

Figure 3: The variance of the angles between item and user embeddings tends to be smaller for tail items than for head items, indicating that tail item embeddings are less effective at capturing user preferences.

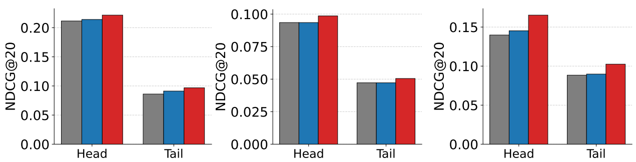
CIKM '24, October 21–25, 2024, Boise, ID, USA.

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/XXXXXX.XXXXXX



(a) Gowalla (b) Yelp (c) MovieLens

Figure 4: The performance of LightGCN, TTEN, and EDGE for the head and tail items, in the unbiased test setting.



(a) Gowalla (b) Yelp (c) MovieLens

Figure 5: The performance of LightGCN, TTEN, and EDGE for the head and tail items, in the biased test setting.

Table 1: The Pearson correlation between embedding magnitudes and popularity of head and tail items.

	Gowalla		Yelp		MovieLens	
	Head	**Tail**	**Head**	**Tail**	**Head**	**Tail**
BPRMF	**0.3328**	0.2240	**0.2915**	0.1293	**-0.5604**	-0.7752
LightGCN	**0.6950**	0.4106	**0.7497**	0.4780	**0.8751**	0.6161

Table 2: The average variance of the angles between item and user embeddings.

	Gowalla		Yelp		MovieLens	
	Head	**Tail**	**Head**	**Tail**	**Head**	**Tail**
BPRMF	**0.0185**	0.0172	**0.0222**	0.0199	**0.0227**	0.0189
LightGCN	**0.0304**	0.0237	**0.0500**	0.0368	**0.0274**	0.0244

B ADDITIONAL EXPERIMENTS

In this section, we provide additional details about the experimental settings and results that were not included in the main paper due to space limitations.

Datasets. In Table 3, we present some statistics of the datasets we used in this paper, Gowalla [1], Yelp [1], and MovieLens [2].

BPR vs. SSM. We present the performance of LightGCN [2] when trained using the BPR loss and SSM loss. Table 4 shows the results in the unbiased test setting, and Table 5 shows the results in the biased setting. In both test settings, the SSM loss is more effective than the BPR loss, which aligns with the finding in [3].

Performance on Head & Tail Items. We evaluate the respective performance of head and tail items. As shown in Table 4 (unbiased test setting) and Table 5 (biased test setting), EDGE improves LightGCN's performance for both head and tail items in unbiased and biased settings.

[1] https://www.yelp.com/dataset
[2] https://grouplens.org/datasets/movielens

Table 3: Summarized statistics of each dataset.

Dataset	# Users	# Items	# Interactions	Density
Gowalla (GW)	29,858	40,981	1,027,370	0.00084
Yelp (YP)	31,668	38,048	1,561,406	0.00130
MovieLens (ML)	69,166	8,790	5,000,415	0.00823

Table 4: The performance (in terms of Recall@20 and NDCG@20) of LightGCN, when trained using the BPR loss and the SSM loss, under the unbiased test setting.

	Gowalla		Yelp		MovieLens	
	Recall	**NDCG**	**Recall**	**NDCG**	**Recall**	**NDCG**
BPR	0.0474	0.0284	0.0069	0.0083	0.0098	0.0064
SSM	**0.0751**	**0.0449**	**0.0173**	**0.0157**	**0.0106**	**0.0067**

Table 5: The performance (in terms of Recall@20 and NDCG@20) of LightGCN, when trained using the BPR loss and the SSM loss, under the biased test setting.

	Gowalla		Yelp		MovieLens	
	Recall	**NDCG**	**Recall**	**NDCG**	**Recall**	**NDCG**
BPR	0.1816	0.1184	0.0909	0.0508	0.3050	0.1570
SSM	**0.2173**	**0.1421**	**0.1155**	**0.0656**	**0.3211**	**0.1634**

REFERENCES

[1] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *KDD*.
[2] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *SIGIR*.
[3] Dain Kim, Jinhyeok Park, and Dongwoo Kim. 2023. Test-Time Embedding Normalization for Popularity Bias Mitigation. In *CIKM*.
[4] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*.
[5] Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. 2023. Deep long-tailed learning: A survey. *TPAMI* (2023).