BaM: An Enhanced Training Scheme for Balanced and Comprehensive Multi-interest Learning

Jaeri Lee
Seoul National University
Seoul, Republic of Korea
jlunits2@snu.ac.kr

Jeongin Yun
Seoul National University
Seoul, Republic of Korea
yji00828@snu.ac.kr

U Kang
Seoul National University
Seoul, Republic of Korea
ukang@snu.ac.kr

ABSTRACT

How can we accurately capture users' diverse interests? To accurately capture users' interests, recent advancements in recommendation systems have led to multi-interest models that use multiple interest vectors. However, practical implementations often rely on a single, well-trained vector, limiting diversity in recommendations. In response, we propose BaM (Balanced Interest Learning for Multi-interest Recommendation), an enhanced training scheme for balanced and comprehensive learning across multiple interest vectors. Unlike traditional methods that focus on the highest similarity vector for loss computation, BaM uses a softer selection criterion, considering multiple relevant vectors. Experiments with real-world datasets show BaM improves accuracy by up to 15.01% in sequential recommendation, achieving state-of-the-art performance.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Multi-interest Recommendation, Sequential Recommendation

ACM Reference Format:
Jaeri Lee, Jeongin Yun, and U Kang. 2024. BaM: An Enhanced Training Scheme for Balanced and Comprehensive Multi-interest Learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), August 25–29, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

How can we accurately capture users' diverse interests? To accurately capture users' diverse interests, recommender systems have evolved significantly with advancements in sequential modeling [6, 8, 9, 12, 14, 16–19, 21, 22, 24–26]. Within these frameworks, multi-interest recommender systems have emerged to address the complex preferences of users by learning and leveraging multiple interest vectors [1–3, 13, 15, 16, 27] This approach allows for a more comprehensive understanding and prediction of user preferences across a broad range of items.

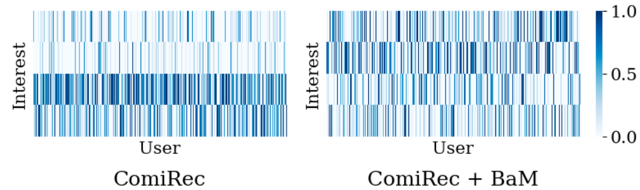

Figure 1: The proportion of items from each interest when the top-10 items are recommended to each user. Darker colors indicate more recommendations from that interest. The heat-map on the left, generated by the multi-interest model ComiRec [1], shows a skewed distribution towards the third and fourth interests. In contrast, applying BaM results in a more balanced distribution among multiple interests, as shown in the heat-map on the right. This approach captures a more comprehensive profile of user preferences, improving recommendation accuracy.

Despite their promise, existing multi-interest recommendation approaches often overlook the importance of balanced training schemes. These models typically prioritize learning a dominant interest representation due to their loss calculation methods, which select a single interest representation based on its similarity to a ground-truth item vector. This bias limits the model's ability to discover and represent less obvious but relevant interests, resulting in skewed recommendations toward dominant interests, as shown in the left heat-map of Figure 1.

To overcome this limitation, we propose a novel training scheme, BaM (Balanced Interest Learning for Multi-interest Recommendation), which employs a softer selection strategy during the training phase. By broadening the scope of interest vectors considered, BaM enhances the model's ability to capture a comprehensive profile of user preferences, leading to more accurate and diverse recommendations. Experimental results demonstrate that applying BaM to models significantly improves the diversity of recommended items, as shown in the right heat-map of Figure 1. BaM achieves up to 15.01% higher accuracy in sequential recommendation compared to the best competitor, resulting in the state-of-the-art performance. The code and dataset are available at https://github.com/jlunits2/BaM.

2 PROPOSED METHOD

In this section, we propose BaM (Balanced Interest Learning for Multi-interest Recommendation), an optimized training approach designed specifically for multi-interest learning models.

To achieve high performance in sequential recommendation with multi-interest learning models, we address the issue of *unbalanced training of multi-interest representation.* It is crucial that each interest representation is adequately trained to capture the wide range

KDD '24, August 25–29, 2024, Barcelona, Spain

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table 1: Summary of datasets.

Dataset	Users	Items	Interactions	Density
Movies & TV [1]	22,747	17,848	841,607	0.20%
Electronics [1]	64,142	31,142	1,475,538	0.07%

[1] https://amazon-reviews-2023.github.io

of users' interests. How can we prevent the training from being biased towards a single interest representation? The main idea of BaM is to use *soft selection of interest*. By softly selecting the interest representation, we aim to distribute training opportunities among multiple interests.

Previous methods of multi-interest recommendation [1–4, 13] offer various techniques to develop a network capable of extracting multi-interest to represent them as matrices. They extract multi-interest $\mathbf{M}_u \in \mathbb{R}^{d\times N}$ for each user given the user's item interaction history using various techniques, where d is the embedding dimension and N is the number of interests. In training stage, an interest representation the most closest to the target item i^+ from the multi-interest representation $\mathbf{M}_u$ is selected:

$$v = \underset{1\leq n\leq N}{\arg\max}(\mathbf{m}_u^{n\top}\mathbf{e}_{i^+}), \tag{1}$$

where v is the index of selected interest representation, $\mathbf{m}_u^n \in \mathbb{R}^d = \mathbf{M}_u[:, n]$ is the n-th interest representation of user u, and $\mathbf{e}_{i^+} \in \mathbb{R}^d$ is the embedding of the target item i^+. However, such hard-selection results in unbalanced training of each interest in the multi-interest representation, because the loss function of the model minimizes the negative log-likelihood using only the selected interest $\mathbf{m}_u^v$:

$$\mathcal{L}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i^+)\in\mathcal{D}} -\log\frac{\exp(\mathbf{m}_u^{v\top}\mathbf{e}_{i^+})}{\sum_{i\in\mathcal{I}}\exp(\mathbf{m}_u^{v\top}\mathbf{e}_i)}, \tag{2}$$

where $\mathcal{D}$ is the training dataset and (u, i^+) is all positive pairs in it.

Instead, we probabilistically select interest vectors using a softmax function to ensure a more balanced training process across multiple interest vectors. By softly selecting the interest representation, we aim to distribute the training opportunities between multiple interests. The probability of user u's n-th interest $\mathbf{m}_u^n$ being selected as the training representation is obtained by applying a softmax function to the correlation score between the multi-interest representation $\mathbf{M}_u$ of user u and the ground-truth item i^+:

$$P(\mathbf{m}_u^n|i^+) = \frac{\exp\left(\mathbf{m}_u^{n\top}\mathbf{e}_{i^+}\right)}{\sum_{1\leq n'\leq N}\exp\left(\mathbf{m}_u^{n'\top}\mathbf{e}_{i^+}\right)}. \tag{3}$$

We calculate the probability for each interest $\mathbf{m}_u^n$ in $\mathbf{M}_u$ and select one according to the probability. The selected interest $\mathbf{m}_u^n$ is used in the loss function instead of $\mathbf{m}_u^v$. This approach can be applied to various multi-interest learning methods without requiring architecture modifications to boost their performance.

3 EXPERIMENTS

We evaluate the performance of BaM compared to competitors in terms of recommendation accuracy.

3.1 Experimental Setup

We introduce our experimental setup including datasets, baselines, and evaluation metrics.

Table 2: Performance of BaM and competitors. BaM shows the best performance in all datasets. n@K and R@K indicate nDCG@K and Recall@K, respectively. The best is marked bold and the second best is underlined.

Dataset	Metric	Method			
		SASRec	ComiRec	REMI	BaM
Movies & TV	n@10	0.0287	0.0201	<u>0.0413</u>	**0.0475**
	n@20	0.0364	0.0251	<u>0.0492</u>	**0.0559**
	R@10	0.0539	0.0372	<u>0.0739</u>	**0.0819**
	R@20	0.0846	0.0572	<u>0.1055</u>	**0.1153**
Books	n@10	0.0492	0.0639	<u>0.0689</u>	**0.0758**
	n@20	0.0629	0.0749	<u>0.0833</u>	**0.0894**
	R@10	0.0961	0.1105	<u>0.1230</u>	**0.1326**
	R@20	0.1505	0.1540	<u>0.1804</u>	**0.1894**

3.1.1 Datasets. We use two real-world datasets as summarized in Table 1. These datasets are from Amazon, one of the world's biggest e-commerce platforms.

3.1.2 Baselines. We compare BaM with single and multi-interest models for sequential recommendation:

- **SASRec** [10] is the state-of-the-art single interest model that utilizes the Transformer [20] architecture's self-attention.
- **ComiRec** [1] is the state-of-the-art multi-interest framework that enables diversity control and incorporates multi-head attention to represent multiple interests.
- **REMI** [23] enhances multi-interest models through an interest-aware hard negative mining strategy and a routing regularization. We apply REMI to ComiRec for the best performance.

3.1.3 Evaluation metrics. We employ leave-one-out protocol [5, 7, 11] where the last item in each user's interactions is removed for testing. Following the prior studies [1, 10, 23], we utilize Recall@K and nDCG@K as metrics of accuracy. We set K to 10 and 20.

3.2 Performance

We compare the performance of BaM and competitors in Table 2. Since BaM is a training scheme, we use ComiRec [1] as the backbone for the best efficiency. We set the embedding dimension as 64 for all models and the number of interests as 4 for all multi-interest models. BaM achieves the best Recall@K and nDCG@K in all datasets showing up to 15.01% improvement compared to the second best method.

4 CONCLUSION

We introduce BaM (<u>Ba</u>lanced Interest Learning for <u>M</u>ulti-interest Recommendation), an advanced training scheme that promotes balanced learning across all interest vectors. Utilizing a softer selection criterion, BaM takes into account a broader range of interest vectors, each representing various aspects of user preferences. This method not only offers a more thorough understanding of user interests but also greatly improves recommendation accuracy. Extensive experiments on real-world datasets demonstrate that BaM surpasses existing models by up to 15.01% in accuracy, achieving state-of-the-art performance.

REFERENCES

[1] Yukuo Cen, Jianwei Zhang, Xu Zou, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Controllable Multi-Interest Framework for Recommendation. In KDD. ACM, 2942–2951.

[2] Zheng Chai, Zhihong Chen, Chenliang Li, Rong Xiao, Houyi Li, Jiawei Wu, Jingxu Chen, and Haihong Tang. 2022. User-Aware Multi-Interest Learning for Candidate Matching in Recommenders. In SIGIR. ACM, 1326–1335.

[3] Gaode Chen, Xinghua Zhang, Yanyan Zhao, Cong Xue, and Ji Xiang. 2021. Exploring Periodicity and Interactivity in Multi-Interest Framework for Sequential Recommendation. In IJCAI. ijcai.org, 1426–1433.

[4] Gaode Chen, Xinghua Zhang, Yanyan Zhao, Cong Xue, and Ji Xiang. 2021. Exploring Periodicity and Interactivity in Multi-Interest Framework for Sequential Recommendation. In IJCAI. ijcai.org, 1426–1433.

[5] Liang Chen, Yang Liu, Xiangnan He, Lianli Gao, and Zibin Zheng. 2019. Matching User with Item Set: Collaborative Bundle Recommendation with Deep Attention Network. In IJCAI.

[6] Ruining He and Julian J. McAuley. 2016. Fusing Similarity Models with Markov Chains for Sparse Sequential Recommendation. In ICDM. IEEE Computer Society, 191–200.

[7] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In WWW. ACM, 173–182.

[8] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2016. Session-based Recommendations with Recurrent Neural Networks. In ICLR (Poster).

[9] Hao Jiang, Wenjie Wang, Yinwei Wei, Zan Gao, Yinglong Wang, and Liqiang Nie. 2020. What Aspect Do You Like: Multi-scale Time-aware User Interest Modeling for Micro-video Recommendation. In ACM Multimedia. ACM, 3487–3495.

[10] Wang-Cheng Kang and Julian J. McAuley. 2018. Self-Attentive Sequential Recommendation. In ICDM. IEEE Computer Society, 197–206.

[11] Jongjin Kim, Hyunsik Jeon, Jaeri Lee, and U Kang. 2023. Diversely Regularized Matrix Factorization for Accurate and Aggregately Diversified Recommendation. In PAKDD.

[12] Yejin Kim, Kwangseob Kim, Chanyoung Park, and Hwanjo Yu. 2019. Sequential and Diverse Recommendation with Long Tail. In IJCAI. ijcai.org, 2740–2746.

[13] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-Interest Network with Dynamic Routing for Recommendation at Tmall. In CIKM. ACM, 2615–2623.

[14] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural Attentive Session-based Recommendation. In CIKM. ACM, 1419–1428.

[15] Yaokun Liu, Xiaowang Zhang, Minghui Zou, and Zhiyong Feng. 2024. Attribute Simulation for Item Embedding Enhancement in Multi-interest Recommendation. In WSDM. ACM, 482–491.

[16] Ruihui Mu. 2018. A Survey of Recommender Systems Based on Deep Learning. IEEE Access (2018).

[17] Aleksandr Vladimirovich Petrov and Craig MacDonald. 2023. gSASRec: Reducing Overconfidence in Sequential Recommendation Trained with Negative Sampling. In RecSys. ACM, 116–128.

[18] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized Markov chains for next-basket recommendation. In WWW. ACM, 811–820.

[19] Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In CIKM. ACM, 1441–1450.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In NIPS. 5998–6008.

[21] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet A. Orgun. 2019. Sequential Recommender Systems: Challenges, Progress and Prospects. In IJCAI. ijcai.org, 6332–6338.

[22] Liwei Wu, Shuqing Li, Cho-Jui Hsieh, and James Sharpnack. 2020. SSE-PT: Sequential Recommendation Via Personalized Transformer. In RecSys. ACM, 328–337.

[23] Yueqi Xie, Jingqi Gao, Peilin Zhou, Qichen Ye, Yining Hua, Jae Boum Kim, Fangzhao Wu, and Sunghun Kim. 2023. Rethinking Multi-Interest Learning for Candidate Matching in Recommender Systems. In RecSys. ACM, 283–293.

[24] YeongWook Yang, Hong-Jun Jang, and Byoungwook Kim. 2020. A Hybrid Recommender System for Sequential Recommendation: Combining Similarity Models With Markov Chains. IEEE Access 8 (2020), 190136–190146.

[25] Yipeng Zhang, Xin Wang, Hong Chen, and Wenwu Zhu. 2023. Adaptive Disentangled Transformer for Sequential Recommendation. In KDD. ACM, 3434–3445.

[26] Kun Zhou, Hui Yu, Wayne Xin Zhao, and Ji-Rong Wen. 2022. Filter-enhanced MLP is All You Need for Sequential Recommendation. In WWW. ACM, 2388–2399.

[27] Zizhong Zhu, Shuang Li, Yaokun Liu, Xiaowang Zhang, Zhiyong Feng, and Yuexian Hou. 2024. High-level preferences as positive examples in contrastive learning for multi-interest sequential recommendation. World Wide Web WWW 27, 2 (2024), 21.