

Train Once, Deploy Anywhere: Matryoshka Representation Learning for Multimodal Recommendation



Yueqi Wang (/profile?id=~Yueqi_Wang2), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Huimin Zeng (/profile?id=~Huimin_Zeng1), Dong Wang (/profile?id=~Dong_Wang21), Julian McAuley (/profile?id=~Julian_McAuley1)

16 Jun 2024 (modified: 02 Aug 2024) ACL ARR 2024 June Submission Everyone Revisions (/revisions?id=atQoQJa3Sz) BibTeX CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

Abstract:

Despite recent advancements in language and vision modeling, integrating rich multimodal knowledge into recommender systems continues to pose significant challenges. This is primarily due to the need for efficient recommendation, which requires adaptive and interactive responses. In this study, we focus on sequential recommendation and introduce a lightweight framework called full-scale Matryoshka representation learning for multimodal recommendation (fMRLRec). Our fMRLRec captures item features at different granularities, learning informative representations for efficient recommendation across multiple dimensions. To integrate item features from diverse modalities, fMRLRec employs a simple mapping to project multimodal item features into an aligned feature space. Additionally, we design an efficient linear transformation that embeds smaller features into larger ones, substantially reducing memory requirements for large-scale training on recommendation data. Combined with improved state space modeling techniques, fMRLRec scales to different dimensions and only requires one-time training to produce multiple models tailored to various granularities. We demonstrate the effectiveness and efficiency of fMRLRec on multiple benchmark datasets, which consistently achieves superior performance over state-of-the-art baseline methods.

Paper Type: Long

Research Area: NLP Applications

Research Area Keywords: multimodal applications

Contribution Types: Approaches to low-resource settings, Approaches low compute settings-efficiency

Languages Studied: English

Author Submission Checklist: I confirm that the paper is anonymous and that all links to data/code repositories in the paper are anonymous., I confirm that the paper has proper length (Short papers: 4 content pages maximum, Long papers: 8 content pages maximum, Ethical considerations and Limitations do not count toward this limit), I confirm that the paper is properly formatted (Templates for *ACL conferences can be found here: <https://github.com/acl-org/acl-style-files> (<https://github.com/acl-org/acl-style-files>).

Association For Computational Linguistics - Blind Submission License Agreement: On behalf of all authors, I agree

Reviewing Volunteers: Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Huimin Zeng (/profile?id=~Huimin_Zeng1)

Reviewing Volunteers For Emergency Reviewing: The volunteers listed above are willing to serve either as regular reviewers or as emergency reviewers.

Reviewing No Volunteers Reason: N/A - An author was provided in the previous question.

TLDR: Parameter efficient multimodal language application for sequential recommendation

Reassignment Request Action Editor: This is not a resubmission

Reassignment Request Reviewers: This is not a resubmission

Preprint: yes

Preprint Status: We are considering releasing a non-anonymous preprint in the next two months (i.e., during the reviewing process).

Preferred Venue: EMNLP 2024

Consent To Share Data: yes

Consent To Share Submission Details: On behalf of all authors, we agree to the terms above to share our submission details.

A1 Limitations Section: This paper has a limitations section.

A2 Potential Risks: Yes

A2 Elaboration: Section 8

A3 Abstract And Introduction Summarize Claims: Yes

A3 Elaboration: Abstract and section 1

B Use Or Create Scientific Artifacts: Yes

B1 Cite Creators Of Artifacts: Yes

B1 Elaboration: Section 5 and Appendix

B2 Discuss The License For Artifacts: Yes

B2 Elaboration: Section 5 and Appendix

B3 Artifact Use Consistent With Intended Use: Yes

B3 Elaboration: Section 5 and Appendix

B4 Data Contains Personally Identifying Info Or Offensive Content: No

B4 Elaboration: Commonly used Publicly Available Datasets without Sensitive Information

B5 Documentation Of Artifacts: Yes

B5 Elaboration: Section 5 and Appendix

B6 Statistics For Data: Yes

B6 Elaboration: Section 5 and Appendix

C Computational Experiments: Yes

C1 Model Size And Budget: Yes

C1 Elaboration: Section 5 and Appendix

C2 Experimental Setup And Hyperparameters: Yes

C2 Elaboration: Section 5 and Appendix

C3 Descriptive Statistics: Yes

C3 Elaboration: Section 6

C4 Parameters For Packages: Yes

C4 Elaboration: Section 5 and Appendix

D Human Subjects Including Annotators: No

D1 Instructions Given To Participants: N/A

D2 Recruitment And Payment: N/A

D3 Data Consent: N/A

D4 Ethics Review Board Approval: N/A

D5 Characteristics Of Annotators: N/A

E Ai Assistants In Research Or Writing: No

E1 Information About Use Of Ai Assistants: N/A

Submission Number: 4681

Discussion (/forum?id=atQoQJa3Sz#discussion)

Filter by reply type Filter by author Search keywords...

Sort: Newest First

Everyone Submission4681... Submission4681 Area... Submission4681 Authors 17 / 17 replies shown

Submission4681... Program Chairs Submission4681... Submission4681...

Submission4681...

Submission4681...



Add:

Author-Editor Confidential Comment

Withdrawal

Meta Review of Submission4681 by Area Chair Fh6G

Meta Review Area Chair Fh6G 07 Aug 2024, 00:31 (modified: 22 Aug 2024, 15:43)

Senior Area Chairs, Area Chairs, Authors, Reviewers Submitted, Program Chairs, Commitment Readers

Revisions (/revisions?id=uQLekR03kU)

Metareview:

This paper proposes a full-scale Matryoshka representation learning framework to learn informative representations for efficient recommendation across multiple dimensions. An efficient linear transformation is constructed to embed smaller features into larger ones, substantially reducing memory requirements for large-scale training on recommendation data. The proposed method can achieve train once and deploy everywhere. The proposed method achieves superior performance over state-of-the-art methods.

Summary Of Reasons To Publish:

1.This paper applies Matryoshka representation learning to the recommendation task and achieves training once and deploying everywhere. Overall, the paper is well organized. 2.The proposed fMRLRec can capture features at different granularities for multimodal recommendation. 3.By designing an efficient linear transformation to embed smaller features into larger ones, memory consumption is significantly reduced.

Summary Of Suggested Revisions:

1. There are some errors in expression that cause confusion and need further correction.
2. It is recommended to further improve the description and explanation of Figure 2 and Formula 14 to facilitate understanding.
3. It is recommended to put the important settings related to the experiments shown in Section 5 instead of the appendix, including other settings such as the structure of image encoder and text encoder.
4. This article needs to further increase the discussion on model interpretability and transparency, which is crucial for the practical application of recommendation systems and gaining user trust.

Overall Assessment: 3 = There are major points that may be revised

Best Paper Ae: No

Ethical Concerns:

There are no concerns with this submission

Needs Ethics Review: No

Author Identity Guess: 1 = I do not have even an educated guess about author identity.

Add:

Author-Editor Confidential Comment

Official Review of Submission4681 by Reviewer KYfj

Official Review Reviewer KYfj 23 Jul 2024, 21:57 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer KYfj, Commitment Readers

Revisions (/revisions?id=4ojeIswRNn)

Paper Summary:

This paper designs a full-scale Matryoshka representation learning to capture features at different granularities for multimodal recommendation (fMRLRec). The proposed fMRLRec can achieve train once and deploy everywhere. The proposed method outperforms state-of-the-art methods with considerable improvements in training efficiency.

Summary Of Strengths:

1. This paper applies Matryoshka representation learning to the recommendation task and achieves training once and deploying everywhere.
2. This paper analyzes the memory efficiency of fMRLRec by driving the number of parameters and activations.

Summary Of Weaknesses:

1. This paper has some similarities with the recent paper "Lai R, Chen L, Chen W, et al. Matryoshka Representation Learning for Recommendation[J]. arXiv preprint arXiv:2406.07432, 2024.", and it is recommended to discuss the similarities and differences with this paper, as well as the advantages.
2. In Eq. 2 " $D(W_i) = D \times kD$ ", it is easy to cause misunderstanding because the D on the left side of the equal sign represents a function name, while the D on the right side of the equal sign represents a scalar.
3. The loss function in Figure 3 is not completely consistent with Eq. 14.
4. Personally, I think the writing in the method section is a bit redundant, and too much space is spent introducing the existing Matryoshka Representation Learning in Sec. 3.2 and Linear Recurrent Units 3.3.2.
5. What kind of image encoder and text encoder framework are used in the proposed method? Is it pre-trained? In order to compare fairly and fully verify the effectiveness of the proposed modules, the compared methods and the proposed method should use the same image and text encoder in the experiment.
6. Aligning visual and textual features is challenging. In Figure 3, a simple mapping layer cannot align textual features with visual features.
7. It seems that fMRLRec can be embedded into most existing multimodal recommendation methods. It is recommended to verify whether fMRLRec can improve the performance of other methods.

Comments Suggestions And Typos:

See weakness

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 3.5

Overall Assessment: 3 = Good: This paper makes a reasonable contribution, and might be of interest for some (broad or narrow) sub-communities, possibly with minor revisions.

Best Paper: No

Needs Ethics Review: No

Reproducibility: 4 = They could mostly reproduce the results, but there may be some variation because of sample variance or minor variations in their interpretation of the protocol or method.

Datasets: 2 = Documentary: The new datasets will be useful to study or replicate the reported research, although for other purposes they may have limited interest or limited usability. (Still a positive rating)

Software: 2 = Documentary: The new software will be useful to study or replicate the reported research, although for other purposes it may have limited interest or limited usability. (Still a positive rating)

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**



Response 1/2 to Reviewer KYfj

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

29 Jul 2024, 06:59 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer KYfj, Commitment Readers

Revisions (/revisions?id=3e1JI2OM02)

Comment:

We are excited that you find our work effective and efficient, we deeply appreciate your constructive comments and hope to clarify your doubts and questions below.

Summary Of Weaknesses:

W1: *This paper has some similarities with the recent paper "Lai R, Chen L, Chen W, et al. Matryoshka Representation Learning for Recommendation..."*

A1: We thank you for bringing up a concurrent work on Matryoshka representation learning [1]. Despite the naming similarity of these two works, we highlight that our fMRLRec learns to compress multimodal item representations for efficient sequential recommendation. By utilizing our efficient weight design and improved linear recurrence, fMRLRec achieves both performance improvements and efficiency compared to state-of-the-art models. In contrast, [1] is based on matrix factorization recommendation and focuses on constructing triplets tailored to capture hierarchical features. In addition, this paper was posted on arxiv on June 11 (4 days before the ARR June submission deadline), and we were not aware of this work during the composing of our submission. We appreciate your mentioning of this paper and will include it as relevant literature along with discussions on the differences between both works. We hope this can address your concerns.

W2: *In Eq. 2 " $D(W_i) = D \times kD$ ", it is easy to cause misunderstanding because the D on the left side...*

A2: We appreciate your suggestions of notation usage, we will substitute the D on the right-hand side of Eq.2 with the lowercase letter d to avoid further confusion.

W3: *The loss function in figure 3 is not completely consistent with Eq. 14.*

A3: We appreciate your detailed feedback on equation 14 and assume what you referred to is actually Figure 2 (which has an MRL loss illustration) instead of Figure 3. We indeed construct the MRL loss function according to Figure 2 by computing multiple loss terms for each model size followed by summation. Specifically, The $\theta[:m]$ denotes the first m elements of the last layer's activation vectors instead of the learned parameters, the activation θ as in Figure 2 is also orthogonal to the time steps $[y_1, y_2, \dots, y_4]$ on the far right-hand side. We will improve the writing by using consistent notations and double-checking on the remaining text to avoid further confusion.

W4: *Personally, I think the writing in the method section is a bit redundant, and too much space...*

A4: We highly value your advice on concise writing and will condense the methodology sections in our revision. However, we kindly point out that section 3.2 actually introduces the proposed full-scale MRL (fMRL), as different from the existing Matryoshka representation learning (MRL) [1], where the former features a weight matrix design that fulfills the purpose of whole-model reusability and improved memory efficiency whereas the latter does not. To the best of our knowledge, the fMRL-like matrix design illustrated in Figure 1 and section 3.2 has not been introduced in previous works. Therefore, fMRL is not a re-adoption or mimicking of existing methods; that is why we detailed the fMRL matrix construction in different cases. As for Linear Recurrent Units (LRU), we give a brief but understandable introduction along with our improvements designed for fMRL (e.g., Pre-LN, SwiGLU, etc.), which already omits certain implementation details (parametrization, actual optimization forms, etc.). Based on the reviewer's suggestion, we will move some pipeline details to the appendix for a better understanding of our paper.

(Response 1/2 finished, continuing in response 2/2)

[1] Lai, Riwei, et al. "Matryoshka Representation Learning for Recommendation." arXiv preprint arXiv:2406.07432 (2024).

[2] Zhai, Xiaohua, et al. "Sigmoid loss for language image pre-training." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[3] Wang, Liang, et al. "Text embeddings by weakly-supervised contrastive pre-training." arXiv preprint arXiv:2212.03533 (2022).



Add: **Author-Editor Confidential Comment**





**Response 2/2
to Reviewer**

KYfj

Official Comment

 Authors ( Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

 29 Jul 2024, 07:07 (modified: 22 Aug 2024, 15:43)

 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer KYfj, Commitment Readers

 Revisions (/revisions?id=3s3nIIsDM0)

Comment:

(Continue from response 1/2)

W5: What kind of image encoder and text encoder framework are used in the proposed method...

A5: We appreciate your question on the details of encoders; we use SigLip [2] for image embedding and E5 [3] for language embedding, which is listed in Appendix A2. As for the second question, yes, both encoder models are pre-trained and frozen during training and inference of our fMRLRec model. We also acknowledge the reviewer's suggestion of using identical encoders for different models. However, baseline methods and our fMRLRec perform image and language encoding in different approaches, and "change encoder" might mean changing the whole model. For instance, RecFormer pretrains a sliding-window-based transformer, and VIP5 integratedly finetunes a P5 model to both encode language and image information. As a result, we found that aligning the language and image encoders among all baseline models was hardly achievable in our experimental setups due to the inherent complexities and limitations encountered. We still maintain fair comparison to the largest extent by using identical datasets, multimodal inputs (same text/image content), and preprocessing strategies for comparison models.

W6: Aligning visual and textual features is challenging. In Figure 3, a simple mapping layer cannot align textual features with visual features.

A6: We appreciate you bringing up this important aspect of modality alignment and respectfully disagree with the reviewer. (1) First, the majority of recent multimodal recommenders and large multimodal models (e.g., MMSSL, VIP5, Llava, Idefics, etc.) adopt a similar approach to project image features, where a pre-trained text/image encoder is used with a linear projector to align representations from different modalities. (2) Empirically, we also observe that a simple linear projector performs the best and improves the training dynamics of fMRLRec, which is consistent with above-mentioned literatures.

W7: It seems that fMRLRec can be embedded into most existing multimodal recommendation...

A7: We thank the reviewer for the valuable suggestion and explain our choice of the LRU backbone here. Since both ID-based and multimodal state-of-the-art recommenders are based on self-attention and its variants, we elaborate on why self-attention models are not used as the backbone of fMRLRec. In self-attention, the hidden states are split into multiple heads to compute softmax attention separately. To integrate fMRL with self-attention, the only option is to split head dimensions into different sizes [8, 16, 32, ...], which differs from the original multi-head design and brings performance losses due to the uneven head dimensions. Additionally, the hidden states in one head can not directly access information from other heads in parallel, which contradicts the design of the proposed fMRL, namely that hidden states of larger dimensions can access more compact hidden states (with the nested representation) to incorporate informative features. Consequently, we selected the recurrence-based SSM models as our backbone, which can be easily integrated with our matrix design and offers improved recommendation efficiency. After empirically evaluating multiple backbones (S4, LRU, Mamba), LRU demonstrated superior performance and we therefore selected LRU as our backbone. We will add explanations on our backbone selection in our updated manuscript and hope this can address the reviewer's concern.

[1] Lai, Riwei, et al. "Matryoshka Representation Learning for Recommendation." arXiv preprint arXiv:2406.07432 (2024).

[2] Zhai, Xiaohua, et al. "Sigmoid loss for language image pre-training." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[3] Wang, Liang, et al. "Text embeddings by weakly-supervised contrastive pre-training." arXiv preprint arXiv:2212.03533 (2022).

Add: **Author-Editor Confidential Comment**



➔ *Replying to Response 2/2 to Reviewer KYfj*

Response to the author's comments

Official Comment Reviewer KYfj 30 Jul 2024, 01:49 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer KYfj, Commitment Readers

Revisions (/revisions?id=AYt6XgaR7O)

Comment:

The author's response addressed most of my concerns, and I revised the score accordingly

Add: **Author-Editor Confidential Comment**



➔ *Replying to Response to the author's comments*

Response to Reviewer KYfj

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

30 Jul 2024, 21:57 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer KYfj, Commitment Readers

Revisions (/revisions?id=HN9eIXhuQS)

Comment:

Thank you so much for considering our rebuttal and for your constructive review on our work! We truly appreciate your recognition and feedback!

Add: **Author-Editor Confidential Comment**



Official Review of Submission4681 by Reviewer no6M

Official Review Reviewer no6M 21 Jul 2024, 01:53 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer no6M, Commitment Readers

Revisions (/revisions?id=MQwtMvwBXX)

Paper Summary:

This paper introduces a novel training framework called fMRLRec for the multimodal sequential recommendation. The primary focus of the study is on addressing the challenges of integrating rich multimodal knowledge into recommender systems efficiently. The framework aims to learn models of varying granularities within a single training procedure, providing an efficient paradigm for multimodal recommendation.

Summary Of Strengths:

- fMRLRec offers an efficient method for training multimodal recommendation systems by embedding smaller features into larger ones. It generates models suitable for various granularities with a single training session, significantly improving training efficiency.
- By designing an efficient linear transformation to embed smaller features into larger ones, memory consumption is significantly reduced.
- The paper provides a detailed theoretical analysis of the model parameters and activation counts of fMRLRec, demonstrating significant savings in parameters and memory usage compared to independently trained models.

Summary Of Weaknesses:

- Although the paper validates the model on multiple benchmark datasets, these datasets are primarily focused on specific domains within Amazon (e.g., Beauty, Clothing, Sports, Toys). It remains unverified whether the model is equally applicable to other types of datasets and application scenarios (e.g., music, movies, news recommendations), which limits the generality of the results.
- The paper provides limited discussion on the interpretability and transparency of the model, which is crucial for the practical application of recommendation systems and for gaining user trust.
- While the paper mentions adopting the LRU recommendation module for fMRLRec, it also highlights the need to test other types of sequential/non-sequential models for a more comprehensive performance evaluation. The lack of detailed comparative analysis with alternative models or state-of-the-art approaches in multimodal recommendation may limit the understanding of how fMRLRec performs relative to existing methods.
- The implementation is not available.

Comments Suggestions And Typos:

- The authors should extend the variety of experimental datasets to include more datasets from different domains and application scenarios (e.g., music recommendation, movie recommendation, news recommendation). This will help validate the effectiveness of the fMRLRec method in a broader range of applications.
- The authors can provide a detailed analysis of the model's resource consumption in practical deployment, along with optimization strategies. This will help readers understand the feasibility of the model in real-world applications.
- For the specific implementation process of the fMRLRec framework and the linear transformations, it is recommended to provide more illustrations and examples. This will help readers understand the complex technical details. Simplifying the descriptions and adding explanatory content will assist readers in better grasping the core ideas of the method.
- Consider exploring the robustness of the fMRLRec framework by conducting sensitivity analyses or investigating the impact of hyperparameters on model performance. Understanding how the framework behaves under different settings and scenarios can provide insights into its stability and reliability in practical applications.

Confidence: 5 = Positive that my evaluation is correct. I read the paper very carefully and am familiar with related work.

Soundness: 3.5

Overall Assessment: 3.5

Best Paper: No

Needs Ethics Review: No

Reproducibility: 3 = They could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined, and/or the training/evaluation data are not widely available.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources



Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**




Discussion Reminder

Official Comment

 Authors ( Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

 30 Jul 2024, 21:59 (modified: 22 Aug 2024, 15:43)

 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer no6M, Commitment Readers

 Revisions (/revisions?id=dcrPYt6Fiv)

Comment:



We thank you for the time and effort you have dedicated to reviewing our paper! We hope our previous responses have addressed your concerns and remain available for any further discussions. Since the discussion period is ending soon, kindly let us know if you have further questions or concerns regarding our work. Thank you!

Add: **Author-Editor Confidential Comment**




Response 2/3 to reviewer no6M

Official Comment

 Authors ( Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

 29 Jul 2024, 13:40 (modified: 22 Aug 2024, 15:43)

 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer no6M, Commitment Readers

 Revisions (/revisions?id=SfNFgk8tTv)

Comment:

(continue from response 1/3)

Comments Suggestions And Typos:

C1: *The authors should extend the variety of experimental datasets...*

A: We appreciate your suggestions on further expanding the datasets selections. Please also refer to our answers to W1 for details.

C2: *The authors can provide a detailed analysis of the model's resource consumption...*

A: We thank you for your suggestion. To the best of our knowledge, widely adopted approaches in real-world recommendation often include multiple stages such as nomination, filtering, re-ranking etc., where models of different granularities are adopted in these stages to balance the performance and efficiency. In addition, item / temporal features are often pre-computed and incrementally used similar to the proposed sequential recommendation framework in fMRLRec. Consequently, training multiple models within a single session can significantly improve the efficiency in learning. Combined with the enhanced training dynamics and autoregressive inference with our improved linear recurrence model, fMRLRec demonstrates substantial improvements in different recommendation scenarios.

To further address the reviewer's concerns, we provide an analysis of memory resources in the table below, where each row represents a different training method and each column a combination of model sizes. All results are obtained on a single NVIDIA A100 (80G) GPU. We compare (1) independent training of different models, and (2) training a fMRLRec model that yields different sizes ready for inference (with no extra cost than training the largest model as introduced in section 3.2). The results demonstrate a significant improvement in the efficiency of fMRLRec than regular, independent task launching. Models of different granularities can be trained within a single session without affecting the model's throughput, thanks to our parameter- and activation-efficient design in fMRLRec.

Model sizes		[1024]	[1024+512]	[1024+... +256]	[1024+... +128]	[1024+... +64]	[1024+... +32]	[1024+... +16]	[1024+... +8]
		1	2	3	4	5	6	7	8
Num. of models		1	2	3	4	5	6	7	8
fMRLRec	Mem. (MB)	7090	7090	7090	7090	7090	7090	7090	7090
Independent	Mem. (MB)	7090	13568	19416	24718	29434	33578	36928	39828
Cost ratio	-	1.00	1.91	2.74	3.49	4.15	4.74	5.21	5.62

We also include an additional throughput analysis during inference time, where we report throughput (TP) as the number of user sequences processed per millisecond (ms) for different model sizes from [8,16,32,...,1024] in the table below. All results are obtained on a single NVIDIA A100 (80G) GPU. The results show that smaller models feature much faster inference speed of multi-thousand-level throughput and are directly extracted from the largest model without extra training cost as introduced in section 3.2.

In summary, fMRLRec's settings of adaptive-sizes favor both low-cost training and efficient inference for users with varying computational resources.

Model size	8	16	32	64	128	256	512	1024
Dataset/TP	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms
Beauty	3727	2484	1597	1064	771	573	438	333
Clothing	6564	4376	2625	1790	1312	984	729	554
Sports	5933	3955	2224	1483	1078	809	613	474
Toys	5099	3399	1912	1274	874	650	493	387

C3: For the specific implementation process of the fMRLRec framework...

A: We appreciate your valuable suggestions on paper writing. We will add additional explanatory descriptions and improve the illustrations in the final version of the paper, our data and implementation will also be released upon publication (also refer to our response to W1).

(Response 2/3 finished, continuing in response 3/3)

Add: **Author-Editor Confidential Comment**



Response 3/3 to reviewer no6M

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

29 Jul 2024, 13:46 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer no6M, Commitment Readers

Revisions (/revisions?id=FNbcBFJgY)

Comment:

(continuing from response 2/3)

C4: Consider exploring the robustness of the fMRLRec framework by conducting...

A: We appreciate your valuable suggestions on analyzing the robustness of our method. Here we provide an additional analysis w.r.t. the selection of r_{\min} and dropout in the following. In our LRU model (recommendation backbone for fMRLRec), r_{\min} is a hyperparameter indicating the initialization magnitude of the decomposed weight matrix A in equation 9. That is, a higher r_{\min} will result in increased historical information in linear recurrence, while a lower r_{\min} indicates higher temporal localness (as shown in Equation 9). We conducted a more thorough search of hyper-parameters for sensitivity analysis thus results in below tables are sometimes higher than reported in paper.

From the first table we observe that our method is quite robust to different selections of r_{\min} , with the optimal selection consistently between 0 and 0.1. The reason for small r_{\min} being optimal (emphasize locality) is that Pretrained multimodal inputs already add rich item context itself, drawing the model's emphasis to the local items and their multimodal representations.

As for dropout shown in the second table, we observe similar robustness and the optimal range can be found between 0.4 to 0.6. The relatively high dropout rate corresponds well to the nature of data sparsity in recommendation datasets [1]. We hope the additional analysis on hyperparameters can fit your suggestions regarding the robustness of our fMRLRec model and thanks again for your advice.

(best metric is in bold, second in italic)

r_{\min}	-	0.0	0.1	0.2	0.3	0.4	0.5
Beauty	N@10	0.0525	<i>0.0520</i>	0.0513	0.0505	0.0495	0.0486
Clothing	N@10	0.0260	<i>0.0259</i>	0.0258	0.0254	0.0253	0.0248
Sports	N@10	0.0278	0.0269	<i>0.0276</i>	0.0263	0.0261	0.0266
Toys	N@10	0.0546	<i>0.0546</i>	0.0540	0.0521	0.0508	0.0497

(best metric is in bold, second in italic)

Dropout	-	0.1	0.2	0.3	0.4	0.5	0.6
Beauty	N@10	0.0497	0.0473	0.0494	0.0515	<i>0.0520</i>	0.0525
Clothing	N@10	0.0245	0.0251	0.0256	0.0262	<i>0.0260</i>	0.0260
Sports	N@10	0.0271	0.0281	0.0282	0.0292	<i>0.0288</i>	0.0278
Toys	N@10	0.0525	0.0538	0.0539	<i>0.0553</i>	0.0558	0.0546

[1] Singh, Monika. "Scalability and sparsity issues in recommender datasets: a survey." Knowledge and Information Systems 62.1 (2020)

(Response 3/3 finished)

Add: **Author-Editor Confidential Comment**



Response 1/3 to reviewer no6M

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

29 Jul 2024, 13:35 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer no6M, Commitment Readers

Revisions (/revisions?id=FQxfGHjYiv)

Comment:

We are excited that you find our work effective and efficient, we deeply appreciate your constructive comments and hope to clarify your doubts and questions below.

Summary Of Weaknesses:

W1: Although the paper validates the model on multiple benchmark datasets...

A1: We appreciate your valuable suggestion and acknowledge that the datasets used in our study are sourced from the Amazon review datasets. The rationale behind this selection is that the Amazon review datasets offer multiple attributes along with images, where the item categories span from beauty products to sport utilities. In contrast, other benchmark datasets predominantly provide only text data or are limited to a single domain (e.g., MovieLens). Therefore, in alignment with previous research (VIP5, MMSSL, etc.), we have chosen to utilize the Amazon review datasets for our experiments. Meanwhile, we are also actively looking for high-quality, multimodal datasets for various domains and hopefully to include one more in final version of our paper.

W2: The paper provides limited discussion on the interpretability and transparency...

A2: Thank you for the insightful feedback. We recognize the importance of interpretability and transparency in the practical application of recommendation systems and for gaining user trust. As a solution, we will address this critical aspect by adding a comprehensive discussion on the interpretability and transparency of our model in our revision. We will also make our code public for improved transparency (also see response for W4 about code availability) and future research.

W3: While the paper mentions adopting the LRU recommendation module for fMRLRec...

A3: We thank the reviewer for the valuable suggestion and explain our choice of the LRU backbone here. Since both ID-based and multimodal state-of-the-art recommenders are based on self-attention and its variants, we elaborate on why self-attention models are not used as the backbone of fMRLRec. In self-attention, the hidden states are split into multiple heads to compute softmax attention separately. To integrate fMRL with self-attention, the only option is to split head dimensions into different sizes [8, 16, 32, ...], which differs from the original multi-head design and brings performance losses due to the uneven head dimensions. Consequently, we selected the recurrence-based SSM models as our backbone, which can be easily integrated with our matrix design and offers improved recommendation efficiency. After empirically evaluating multiple backbones (S4, LRU, Mamba), LRU demonstrated superior performance and we therefore selected LRU as our backbone. We will add explanations on our backbone selection in our updated manuscript and hope this can address the reviewer's concern.

W4: The implementation is not available.

A4: We thank you for your suggestions, we are more than happy to make public all of the adopted datasets and implementation to boost future research. However, since ARR policies do not allow links at this stage of discussion, we here provide PyTorch code of fMRLRec's core functions (annotated) to show it is easy-to-implement, code are directly copied from our executables:

For creating fMRLRec masks introduced in section 3.2, case 1 (also shown in Figure 1):

```

import torch
def create_fmrl_weight_mask_upscale(up_size, fmrl_sizes):
    """
    Args:
        up_size: the resulting dimension after upscale.
        fmrl_sizes: the list of fMRL sizes e.g. [8,16,32...,1024]
    Output:
        fMRL-based weight mask for the upscale weight
    """
    # compute upscale multiple
    k = int(up_size/fmrl_sizes[-1])
    mask = torch.zeros(fmrl_sizes[-1], up_size) # (d, 2d)
    # iterate through each fMRL size
    for i in range(len(fmrl_sizes)):
        if i == 0:
            # create masking for the first fMRL size (section 3.2 case 1)
            mask[0:fmrl_sizes[0], 0:k*fmrl_sizes[0]]=1
        else:
            # create masking for the remaining fMRL sizes (section 3.2 case 1)
            mask[0:fmrl_sizes[i], k*fmrl_sizes[i-1]:k*fmrl_sizes[i]]=1
    # transpose to suit pytorch weight sizes
    return mask.transpose(0,1)

```

For applying fMRL mask to a model weight in Linear recurrent units or FFN, and exerting the masked weight on input x of size (batch size, sequence length, dimension of the model) as introduced in section 3.2, Eq. 5:

```

self.w_1_mask = create_fmrl_weight_mask_upscale(self.w_1.weight.shape[0], self.args.mrl_hidden_sizes)
x_ = nn.functional.linear(x, self.w_1.weight * self.w_1_mask, self.w_1.bias)

```

For computing dot-product based logits between model final output x and embedding weight for fMRLRec Cross-Entropy loss (Section 3.3.4, Eq. 13 and Eq. 14):

```



logits = torch.cat([torch.matmul(x[...,:s], embedding_weight[...,:s].permute(1,0)) + self.bias \
                    for _, s in enumerate(self.args.mrl_hidden_sizes)], dim=0)
loss = nn.CrossEntropyLoss()(logits, labels)

```

(continuing)

Add: Author-Editor Confidential Comment

Official Review of Submission4681 by Reviewer tEbn

Official Review  Reviewer tEbn  13 Jul 2024, 19:57 (modified: 22 Aug 2024, 15:43)

 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

 Revisions (/revisions?id=13Uunscaik)

Paper Summary:

The paper introduces fMRLRec, a novel framework designed to tackle the challenge of integrating multimodal knowledge into recommender systems efficiently. fMRLRec focuses on sequential recommendation and is built to handle different granularities of item features, learning informative representations that can be deployed across multiple dimensions

without requiring additional training sessions. The framework uses a simple mapping to align multimodal features and applies efficient linear transformations to manage memory requirements during large-scale training. It demonstrates improved performance compared to state-of-the-art methods through comprehensive benchmark testing.

Summary Of Strengths:

- well-written
- this paper is easy to follow

Summary Of Weaknesses:

- Not sure if the author will open source the code for this work. It is difficult to assess the reproducibility of the work in this paper, and I am at the same time concerned about the reliability of the experimental results in this paper
- The paper lacks a thorough (at least preliminary) theoretical explanation for why fMRLRec is effective in recommender systems. The overall lack of novelty in the proposed approach makes it difficult to inspire communities in a substantive way
- Despite the claim of efficiency, the computational resources required for training multimodal models, especially when dealing with large datasets, are still substantial. The paper does not delve deeply into how fMRLRec addresses this issue

Comments Suggestions And Typos:

see weakness

Confidence: 4 = Quite sure. I tried to check the important points carefully. It's unlikely, though conceivable, that I missed something that should affect my ratings.

Soundness: 3 = Acceptable: This study provides sufficient support for its major claims/arguments. Some minor points may need extra support or details.

Overall Assessment: 2.5

Best Paper: No

Needs Ethics Review: No

Reproducibility: 3 = They could reproduce the results with some difficulty. The settings of parameters are underspecified or subjectively determined, and/or the training/evaluation data are not widely available.

Datasets: 1 = No usable datasets submitted.

Software: 1 = No usable software released.

Knowledge Of Or Educated Guess At Author Identity: No

Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Knowledge Of Paper Source: N/A, I do not know anything about the paper from outside sources

Impact Of Knowledge Of Paper: N/A, I do not know anything about the paper from outside sources

Add: **Author-Editor Confidential Comment**



Discussion Period Ending Soon

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

01 Aug 2024, 23:51 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

Revisions (/revisions?id=TiyFZfjE6y)

Comment:

Again, we thank you for the time and efforts you have dedicated to reviewing our work! As the discussion period is ending today, we kindly remind you to acknowledge our response to your reviews. We hope our previous responses could address your concerns and are happy to follow-up in further discussion if any concerns remain. Please find our previous responses to your feedback below.

Add: **Author-Editor Confidential Comment**



Discussion Reminder

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

31 Jul 2024, 23:28 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

Revisions (/revisions?id=piO4Yv5wQU)

Comment:

We thank you for the time and effort you have dedicated to reviewing our fMRLRec! We hope our previous responses have addressed your concerns and remain available for any further discussions. Since the discussion period is ending, kindly let us know if you have further questions or concerns regarding our work. Thank you!

Add: **Author-Editor Confidential Comment**



Discussion Reminder

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

30 Jul 2024, 21:58 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

Revisions (/revisions?id=IkZoPKQYhe)

Comment:

We appreciate your constructive feedback in the review! We hope our previous responses could address your concerns on our method. Since the discussion period is ending soon, please let us know if you have any additional questions or concerns. Thank you!

Add: **Author-Editor Confidential Comment**



Response 2/2 to reviewer tEbn

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=aclweb.org/ACL/ARR/2024/June/Submission4681/Authors))

29 Jul 2024, 10:37 (modified: 22 Aug 2024, 15:43)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

Revisions (/revisions?id=5YN37158bZ)

Comment:

(continue from response 1/2)

W3: *Despite the claim of efficiency, the computational resources required for training multimodal models...*

A3: We thank you for your suggestion. To the best of our knowledge, widely adopted approaches in real-world recommendation often include multiple stages such as nomination, filtering, re-ranking etc., where models of different granularities are adopted in these stages to balance the performance and efficiency. In addition, item / temporal features are often pre-computed and incrementally used similar to the proposed sequential recommendation framework in fMRLRec. Consequently, training multiple models within a single session can

significantly improve the efficiency in learning. Combined with the enhanced training dynamics and autoregressive inference with our improved linear recurrence model, fMRLRec demonstrates substantial improvements in different recommendation scenarios.

To further address the reviewer's concerns, we provide an analysis of memory resources in the table below, where each row represents a different training method and each column a combination of model sizes. All results are obtained on a single NVIDIA A100 (80G) GPU. We compare (1) independent training of different models, and (2) training a fMRLRec model that yields different sizes ready for inference (with no extra cost than training the largest model as introduced in section 3.2). The results demonstrate a significant improvement in the efficiency of fMRLRec than regular, independent task launching. Models of different granularities can be trained within a single session without affecting the model's throughput, thanks to our parameter- and activation-efficient design in fMRLRec.

Model sizes		[1024]	[1024+512]	[1024+...+256]	[1024+...+128]	[1024+...+64]	[1024+...+32]	[1024+...+16]	[1024+...+8]
Num. of models		1	2	3	4	5	6	7	8
fMRLRec	Mem. (MB)	7090	7090	7090	7090	7090	7090	7090	7090
Independent	Mem. (MB)	7090	13568	19416	24718	29434	33578	36928	39828
Cost ratio	-	1.00	1.91	2.74	3.49	4.15	4.74	5.21	5.62

We also include an additional throughput analysis during inference time, where we report throughput (TP) as the number of user sequences processed per millisecond (ms) for different model sizes from [8,16,32,...,1024] in the table below. All results are obtained on a single NVIDIA A100 (80G) GPU. The results show that smaller models feature much faster inference speed of multi-thousand-level throughput and are directly extracted from the largest model without extra training cost as introduced in section 3.2.

In summary, fMRLRec's settings of adaptive-sizes favor both low-cost training and efficient inference for users with varying computational resources.

Model size	8	16	32	64	128	256	512	1024
Dataset/TP	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms	users/ms
Beauty	3727	2484	1597	1064	771	573	438	333
Clothing	6564	4376	2625	1790	1312	984	729	554
Sports	5933	3955	2224	1483	1078	809	613	474
Toys	5099	3399	1912	1274	874	650	493	387

Comments Suggestions And Typos:

Please refer to weakness.

Add: **Author-Editor Confidential Comment**



Response 1/2 to reviewer tEbn

Official Comment

Authors (Yueqi Wang (/profile?id=~Yueqi_Wang2), Huimin Zeng (/profile?id=~Huimin_Zeng1), Zhenrui Yue (/profile?id=~Zhenrui_Yue1), Dong Wang (/profile?id=~Dong_Wang21), +1 more (/group/info?id=acweb.org/ACL/ARR/2024/June/Submission4681/Authors))

29 Jul 2024, 10:21 (modified: 22 Aug 2024, 15:43)

👁 Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Authors, Reviewer tEbn, Commitment Readers

📄 Revisions (/revisions?id=FAwobLSkjO)

Comment:

We thank you for your constructive comments on our work and hope to clarify your doubts and questions in the following.

Summary Of Weaknesses

W1: *Not sure if the author will open source the code for this work. It is difficult to ...*

A1: We thank you for your suggestions, we are more than happy to make public all of the adopted datasets and implementation to boost future research. However, since ARR policies do not allow links at this stage of discussion, we here provide PyTorch code of some core functions of fMRLRec design with annotations to show it is simple and easy-to-implement, code snippets are directly copied from our executables:

For creating fMRLRec masks as introduced in section 3.2, case 1 (also illustrated in Figure 1):

```
import torch
def create_mrl_weight_mask_upscale(up_size, fmrl_sizes):
    """
    Args:
        up_size: the resulting dimension after upscale.
        fmrl_sizes: the list of fMRL sizes e.g. [8,16,32...,1024]
    Output:
        fMRL-based weight mask for the upscale weight
    """
    # compute upscale multiple
    k = int(up_size/fmrl_sizes[-1])
    mask = torch.zeros(fmrl_sizes[-1], up_size) # (d, 2d)
    # iterate through each fMRL size
    for i in range(len(fmrl_sizes)):
        if i == 0:
            # create masking for the first fMRL size (section 3.2 case 1)
            mask[0:fmrl_sizes[0], 0:k*fmrl_sizes[0]]=1
        else:
            # create masking for the remaining fMRL sizes (section 3.2 case 1)
            mask[0:fmrl_sizes[i], k*fmrl_sizes[i-1]:k*fmrl_sizes[i]]=1
    # transpose to suit pytorch weight sizes
    return mask.transpose(0,1)
```

For applying fMRL mask to a model weight in Linear recurrent units or FFN, and exerting the masked weight on input x of size (batch size, sequence length, dimension of the model) as introduced in section 3.2, Eq. 5:

```
self.w_1_mask = create_mrl_weight_mask_upscale(self.w_1.weight.shape[0], self.args.mrl_hidden_sizes)
x_ = nn.functional.linear(x, self.w_1.weight * self.w_1_mask, self.w_1.bias)
```

For computing dot-product based logits between model final output x and embedding weight for fMRLRec Cross-Entropy loss (Section 3.3.4, Eq. 13 and Eq. 14):

```
logits = torch.cat([torch.matmul(x[...,:s], embedding_weight[...,:s].permute(1,0)) + self.bias \
                    for _, s in enumerate(self.args.mrl_hidden_sizes)], dim=0)
loss = nn.CrossEntropyLoss()(logits, labels)
```

W2: *The paper lacks a thorough (at least preliminary) theoretical explanation...*

A2: Thank you for your feedback, we acknowledge that theoretical analysis could complement the performance and efficiency of fMRL. We here provide a preliminary analysis of why fMRLRec is an effective multimodal recommender by making below points:

(i) Recommendation datasets in general feature a strong data sparsity [1] as a majority of products (e-commerce) or movies (for movie recommendation) in the pool are not bought/seen by a user, and vice-versa for the products' relationship to users.

(ii) This natural sparsity easily causes overfitting [1] thus it favors small model sizes around (64,128,256) for classic, non-multimodal recommenders [2], which contradicts the relatively larger model sizes required for multimodal encodings (SigLip of size 768 [3], E5 large of size 1024 [4]).

(iii) Our fMRLRec model features a joint learning of all model sizes (Eq. 14) where gradients from explicit loss terms of different granularities/model sizes (Figure 2) are jointly utilized to optimize all model sizes, leveraging information from all dimensions to mitigate the above-mentioned size contradiction. This optimization choice, further paired with our fMRL weight design, allows fMRLRec to achieve state-of-the-art performance compared to strong multimodal baselines.

To the best of our knowledge, a strict mathematical analysis of fMRLRec is yet to surface and we would like to include it in the future work. We would also respectfully suggest that the main focus of this paper is practical effectiveness and memory efficiency (also refer to response 3 for additional experiments). We will include this above analysis into the final version to pair with our empirical results for better comprehension.

(Response 1/2 finished, continuing in response 2/2)

[1] Singh, Monika. "Scalability and sparsity issues in recommender datasets: a survey." Knowledge and Information Systems 62.1 (2020)

[2] Fang, Hui, et al. "Deep learning for sequential recommendation: Algorithms, influential factors, and evaluations." ACM Transactions on Information Systems (TOIS) 39.1 (2020): 1-42.

[3] Zhai, Xiaohua, et al. "Sigmoid loss for language image pre-training." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[4] Wang, Liang, et al. "Text embeddings by weakly-supervised contrastive pre-training." arXiv preprint arXiv:2212.03533 (2022).

Add: **Author-Editor Confidential Comment**

[About OpenReview \(/about\)](/about)

[Hosting a Venue \(/group?id=OpenReview.net/Support\)](/group?id=OpenReview.net/Support)

[All Venues \(/venues\)](/venues)

[Sponsors \(/sponsors\)](/sponsors)

[Frequently Asked Questions](#)

[\(https://docs.openreview.net/getting-started/frequently-asked-questions\)](https://docs.openreview.net/getting-started/frequently-asked-questions)

[Contact \(/contact\)](/contact)

[Feedback](#)

[Terms of Use \(/legal/terms\)](/legal/terms)

[Privacy Policy \(/legal/privacy\)](/legal/privacy)

[OpenReview \(/about\)](/about) is a long-term project to advance science through improved peer review, with legal nonprofit status through [Code for Science & Society \(https://codeforscience.org/\)](https://codeforscience.org/). We gratefully acknowledge the support of the [OpenReview Sponsors \(/sponsors\)](/sponsors). © 2024 OpenReview