Fusion and Alignment Enhancement with Large Language Models for Tail-item Sequential Recommendation

Zhifu Wei
Northeastern University
Shenyang, China
weizf@mails.neu.edu.cn

Yizhou Dang
Northeastern University
Shenyang, China
dangyz@mails.neu.edu.cn

Guibing Guo*
Northeastern University
Shenyang, China
guogb@swc.neu.edu.cn

Chuang Zhao
The Hong Kong University of Science and Technology
Hong Kong, China
czhaobo@connect.ust.hk

Zhu Sun*
Singapore University of Technology and Design
Singapore, Singapore
sunzhuntu@gmail.com

Abstract

Sequential Recommendation (SR) learns user preferences from their historical interaction sequences and provides personalized suggestions. In real-world scenarios, most items exhibit sparse interactions, known as the tail-item problem. This issue limits the model's ability to accurately capture item transition patterns. To tackle this, large language models (LLMs) offer a promising solution by capturing semantic relationships between items. Despite previous efforts to leverage LLM-derived embeddings for enriching tail items, they still face the following limitations: 1) They struggle to effectively fuse collaborative signals with semantic knowledge, leading to suboptimal item embedding quality. 2) Existing methods overlook the structural inconsistency between the ID and LLM embedding spaces, causing conflicting signals that degrade recommendation accuracy. In this work, we propose a **F**usion and **A**lignment **E**nhancement framework with LLMs for Tail-item Sequential **Rec**ommendation (**FAERec**), which improves item representations by generating coherently-fused and structurally consistent embeddings. For the information fusion challenge, we design an adaptive gating mechanism that dynamically fuses ID and LLM embeddings. Then, we propose a dual-level alignment approach to mitigate structural inconsistency. The item-level alignment establishes correspondences between ID and LLM embeddings of the same item through contrastive learning, while the feature-level alignment constrains the correlation patterns between corresponding dimensions across the two embedding spaces. Furthermore, the weights of the two alignments are adjusted by a curriculum learning scheduler to avoid premature optimization of the complex feature-level objective. Extensive experiments across three widely used datasets with multiple representative SR backbones demonstrate the effectiveness and generalizability of our framework. The codes are provided at https://github.com/ZhifuWei/FAERec.

*Corresponding authors.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
SIGIR '26, Melbourne, VIC, Australia
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2599-9/2026/07
https://doi.org/10.1145/3805712.3809741

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Sequential Recommendation; Long-Tail Problem; Large Language Models; Fusion and Alignment

ACM Reference Format:
Zhifu Wei, Yizhou Dang, Guibing Guo, Chuang Zhao, and Zhu Sun. 2026. Fusion and Alignment Enhancement with Large Language Models for Tail-item Sequential Recommendation. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '26), July 20–24, 2026, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3805712.3809741

1 Introduction

Sequential recommendation aims to predict the next items of interest to users based on their historical interaction behavior and has been widely applied across various real-world scenarios such as e-commerce, streaming media, and social networks [25, 49, 52]. Learning high-quality item embeddings is essential for models to effectively capture item transition patterns and users' behaviors [11, 12, 50]. Early research learned ID embeddings for each item, capturing collaborative signals through user-item interaction data [24, 31].

Despite significant advances in sequential recommendation, the tail-item problem remains a critical challenge, undermining item embedding quality and user experience [7, 29, 33]. As shown in Figure 1 (a), we trained SASRec [25], a representative sequential recommendation model, on the Amazon Beauty dataset and display its performance across items grouped by interaction frequency. The histogram reveals that most items have fewer than 5 records, while the corresponding line graph indicates their relatively low performance. This stems from an inherent limitation of ID-based models, which rely solely on collaborative signals from interaction data. For tail items with sparse interactions, such models fail to provide adequate information for effective representation learning [24]. In contrast, abundant interactions enable head items to leverage rich collaborative patterns, resulting in significantly better performance.

To tackle the tail-item challenge, recent advances in large language models (LLMs) offer new opportunities by capturing semantic item relationships from textual attributes [38, 40, 41, 51]. Motivated

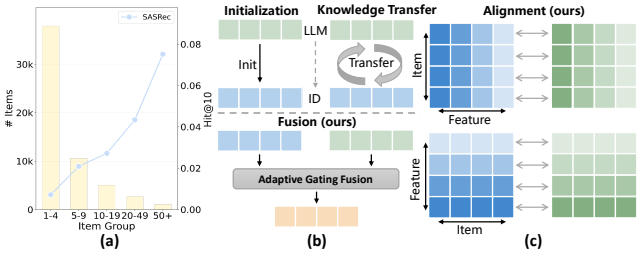


Figure 1: Visualization of the long-tail item problem (a) and comparison between the existing LLM-based sequential recommendation methods and our FAERec (b, c).

by this, several semantic-guided learning strategies have been proposed to enhance ID embeddings [21, 38, 39, 46, 65]. 1) **Semantic initialization** [18, 46, 56] utilizes language embeddings to provide effective initialization for the ID embedding layer, gradually evolving from semantic space to collaborative space. 2) **Knowledge transfer** [38, 47] maximizes mutual information between ID and LLM embeddings through cross-view contrastive learning.

Despite their effectiveness, these methods face two critical challenges: (i) **Insufficient Semantic Leveraging.** As shown in Figure 1 (b), semantic initialization and knowledge transfer strategies use LLM embeddings solely as auxiliary signals for learning ID embeddings, without explicitly retaining LLM embeddings in the final item embeddings, leading to the suboptimal item embedding quality. (ii) **Inconsistent Space Structures.** While achieving cross-view knowledge transfer, existing methods overlook the structural alignment between the ID and LLM embedding spaces. Specifically, the embedding space structure is determined by how each dimension encodes item information and how dimensions correlate with each other [13, 36, 64]. Since ID embeddings capture collaborative patterns while LLM embeddings encode semantic textual attributes, the two spaces organize information through inconsistent dimensional structures [43]. This results in inconsistent item neighborhoods across the two embedding spaces: Items close in collaborative space may be distant in semantic space (*due to space limitations, this is verified in Section 5.6*). Such structural inconsistency introduces conflicting signals when leveraging LLM semantic knowledge, thereby degrading recommendation accuracy.

In this paper, we propose a **F**usion and **A**lignment **E**nhancement framework with LLMs for Tail-item Sequential **Rec**ommendation (**FAERec**). Our core idea is to improve semantic utilization and structural consistency by effectively fusing and aligning ID and LLM embeddings. Specifically, we first derive semantic embeddings by encoding item texts with LLMs, which can be cached to avoid extra inference costs. To further leverage semantic information, we perform dimension-wise weighted fusion of ID and LLM embeddings, allowing the model to capture fine-grained complementary patterns between collaborative and semantic signals. The fusion weights are adaptively determined by a gating mechanism based on the characteristics of each item's embeddings. For structural inconsistency, as shown in Figure 1 (c), we design a dual-level alignment approach. The item-level alignment utilizes contrastive learning to maximize the similarity between ID and LLM embeddings of the same item while minimizing it for different items, establishing point-to-point correspondences across the two embeddings. At the feature level, inspired by the Barlow Twins [64], we first standardize the ID and LLM embeddings separately and compute their cross-correlation matrix. Furthermore, we constrain this matrix to align corresponding dimensions while decoupling different dimensions, achieving structural consistency across embedding spaces. To prevent premature convergence to the complex feature-level objective, we introduce a curriculum learning scheduler based on cosine annealing, balancing the two alignment objectives. Extensive experiments on various datasets and backbones demonstrate that our proposed method achieves significant improvements.

In summary, our work makes the following contributions:

- We highlight insufficient semantic leveraging and inconsistent space structures of existing LLM-based SR methods and propose a **F**usion and **A**lignment **E**nhancement framework with LLMs for Sequential **Rec**ommendation, which can improve the quality of item embeddings and alleviate the tail-item challenge.
- We design two key modules. Adaptive gating fusion effectively integrates collaborative signals and semantic knowledge. Dual-level alignment establishes item-level correspondences between ID and LLM embeddings and feature-level structural consistency across the two embedding spaces, guided by curriculum learning.
- We conduct comprehensive experiments on real-world datasets with representative backbone SR models and different baselines to demonstrate the effectiveness and generalizability of FAERec.

2 RELATED WORK

2.1 General Sequential Recommendation

Sequential Recommendation (SR) aims to predict users' next item of interest based on their interaction sequences [10, 67, 69]. Early research leveraged Markov Chains [19] to extract user preference, focusing on correlations between the next item and recent interactions. Later, deep learning architectures such as CNNs [62] and RNNs [45] were adopted: GRU4Rec [20] used gated recurrent units to model interest evolution, while Caser [52] projected items to latent spaces and learned patterns via convolution operations. With the advancement of self-attention [54, 66], Transformer-based SR models gained prominence. SASRec [25] adopted the multi-head attention mechanism to characterize transition correlations effectively. BERT4Rec [48] introduced a two-directional attention encoder trained with a cloze-style prediction task. Departing from Transformer-based methods, research showed that filtering mechanisms effectively mitigate noise [72], which motivated the proposal of an all-MLP framework with trainable filters. To efficiently learn sequential transitions, LRURec [63] introduced matrix diagonalization-based linear recurrence, complemented by a recursive parallelization scheme that improves training speed.

2.2 Long-tail Sequential Recommendation

To alleviate the tail-item problem, early research focuses on leveraging collaborative signals from head items with abundant interactions to enrich tail-item representations [24, 26]. As a representative work, CITIES [24] leverages multiple contextual items from the head to train an embedding inference function, improving the embedding quality of tail items. Inspired by this idea, MELT [26] designs a mutually reinforcing dual-branch framework to enhance

performance for both tail users and tail items by transferring knowledge from head users and items, respectively. MASR [23] proposes centroid-wise and cluster-wise memory banks to store historical interaction patterns and retrieve similar sequences for tail items. LOAM [61] generates augmented interaction sequences that end with tail items through global context integration, and leverages cross-session mixup to synthesize new interactions to enrich tail item representations. Despite their effectiveness, these approaches fundamentally rely on collaborative signals extracted from interaction data. Tail items with highly sparse interactions fail to provide sufficient contextual information for these methods to learn effective representations, resulting in suboptimal item embeddings.

2.3 LLMs for Recommendation

Large language models have attracted widespread attention due to their powerful abilities in semantic understanding [37, 68, 70, 71]. Recent research has explored integrating LLMs into recommender systems, which can be categorized into two primary approaches [21, 32, 38, 46, 58, 59]. The first approach directly employs LLMs for recommendation tasks. For example, TALLRec [2] uses instruction tuning to enable LLMs to generate recommendations. ChatRec [15] designs prompt templates to activate LLM capabilities. The second approach leverages LLM embeddings to enhance traditional models. Semantic initialization methods, such as LLM2Bert4Rec [18] and SAID [22], utilize LLM embeddings to initialize ID embeddings, which subsequently adapt to collaborative patterns during training. RLMRec [47] proposes knowledge transfer through mutual information maximization, establishing bidirectional mappings between language and behavior spaces. LLM-ESR [39] introduces dual-view modeling and retrieval-augmented distillation to improve representations for both tail users and items in sequential recommendation. Unlike the above efforts, our proposed FAERec is tailored from both the fusion and alignment perspectives, furthermore advancing LLM-driven item representation learning for tail items.

3 PRELIMINARIES

Problem Definition. Suppose we have a set of users $\mathcal{U}$ and a set of items $\mathcal{V}$. Each user $u \in \mathcal{U}$ corresponds to a chronologically ordered sequence of interacted items $s_u = [v_1, ..., v_j, ..., v_{|s_u|}]$, where $v_j \in \mathcal{V}$ denotes the item u interacted with at time step j, and $|s_u|$ is the sequence length. Given a user's interaction item sequence s_u, the sequence recommendation objective is to accurately predict the item v^* most likely to be interacted with by user u at time step $|s_u| + 1$. Its mathematical expression is:

$$\arg\max_{v^* \in \mathcal{V}} P\left(v_{|s_u|+1} = v^* \mid s_u\right) \tag{1}$$

Head/Tail Splitting. Referring to existing research on long-tail item sequential recommendations [26, 39], we categorize items into tail and head groups. Let p_v denote the popularity of item v. We sort items in descending order based on p_v. Subsequently, following the Pareto principle [3], we select the top 20% of items as head items ($\mathcal{V}_{\text{head}}$), with the remaining items constituting the tail items, i.e., $\mathcal{V}_{\text{tail}} = \mathcal{V} \setminus \mathcal{V}_{\text{head}}$ [39]. To mitigate the tail-item challenge, we aim to enhance recommendation performance specifically for $\mathcal{V}_{\text{tail}}$.

4 OUR METHOD

In this section, we present our proposed FAERec. The Overview framework is illustrated in Figure 2. Our approach consists of two key modules: adaptive gating fusion and dual-level alignment. We then introduce the training and inference processes of FAERec. Finally, we discuss our FAERec and existing methods.

4.1 Adaptive Gating Fusion

Traditional SR models use randomly initialized ID embeddings to learn collaborative patterns, achieving strong performance on popular items but struggling with tail items due to sparse interactions [26, 33]. To address this, we enhance item embeddings by fusing collaborative signals from ID embeddings with semantic knowledge from LLM embeddings. Besides, we introduce a dimension-wise weighting strategy to combine the benefits of both better.

ID Embedding Layer. For each item $v_i \in \mathcal{V}$, we maintain a trainable embedding vector $\mathbf{e}_i^{ID} \in \mathbb{R}^d$, forming the ID embedding matrix $\mathbf{E}^{ID} \in \mathbb{R}^{|\mathcal{V}| \times d}$. These embeddings are updated during model training via backpropagation to capture underlying item transition patterns and collaborative signals. Given an original user sequence s_u, the embedding Look-up operation will be applied for E^{ID} to obtain an item embedding sequence, i.e, $E_u^{ID} = [\mathbf{e}_1^{ID}, \mathbf{e}_2^{ID}, \ldots, \mathbf{e}_{|s_u|}^{ID}]$.

LLM Embedding Layer. Typically, the attributes and descriptions of items contain rich semantic information. To leverage the text comprehension capabilities of LLMs, we organize item attributes (e.g., title, category and description) into natural language prompts and encode them using a pre-trained LLM to obtain LLM embeddings [47]. Specifically, we construct the following item prompts:

Item Prompt Template (Beauty)

The Beauty item has the following attributes:
name is <TITLE>; brand is <BRAND>; rating is <DATE>; price is <PRICE>.
The item has following features: <CATEGORIES>.
The item has following descriptions: <DESCRIPTION>.

By substituting the template fields with corresponding attribute values, item prompts are input to LLMs to extract rich semantic embeddings. In practice, LLM embeddings can be obtained through open-source models like LLaMA [53] or commercial APIs such as OpenAI's text-embedding-ada-002 [1]. We employ the latter to ensure high-quality and stable embeddings. Let $\mathbf{E}^{LLM} \in \mathbb{R}^{d_{llm}}$ denote the LLM embeddings of all items, where d_{llm} is the embedding dimension. All item LLM embeddings are generated offline in advance and cached before training, thereby avoiding the computational overhead of online inference. Besides, since the dimension d_{LLM} of original LLM embeddings (e.g., 1536 dimensions) is significantly higher than the hidden dimension d of the recommendation model (typically 64 or 128), we designed a lightweight projection network to perform dimension adaptation[22]. Specifically, we obtain the original LLM embedding $\mathbf{e}_{ori}^{LLM}$ from the LLM API, and project it to match the ID embedding dimension:

$$\mathbf{e}_i^{LLM} = f_\theta(\mathbf{e}_{ori}^{LLM}) \tag{2}$$

[1] https://platform.openai.com/

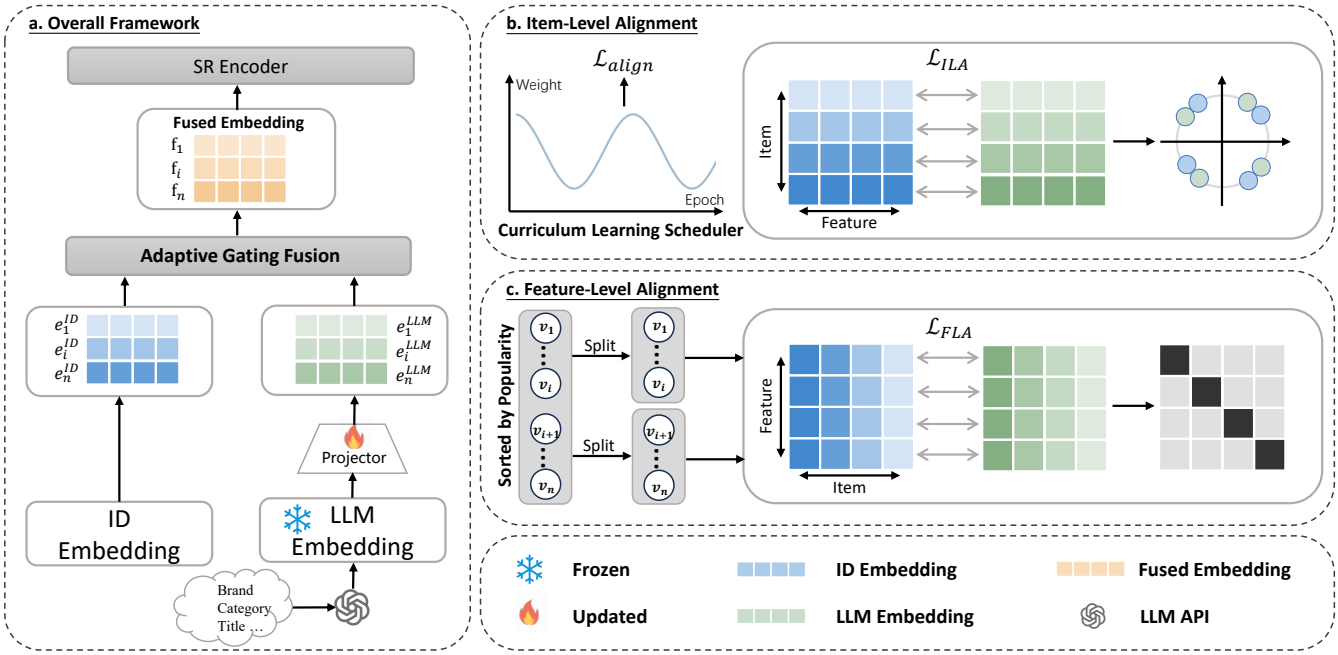


Figure 2: The overall framework of our proposed FAERec.

where $f_\theta(\cdot)$ is implemented via two layers of MLP. Then, we obtain the LLM embeddings corresponding to user u's interaction sequence, denoted as $\mathbf{E}_u^{LLM} = [\mathbf{e}_1^{LLM}, \mathbf{e}_2^{LLM}, \ldots, \mathbf{e}_{|s_u|}^{LLM}]$. The projection network parameters are trainable, while the original LLM embeddings remain frozen throughout training . Besides, since LLM embeddings are obtained from pretrained models while ID embeddings are learned from scratch, directly combining them may lead to an imbalance in the optimization [1, 39]. To mitigate this, we use Principal Component Analysis (PCA) to reduce LLM embeddings to d dimensions and use them to initialize ID embeddings.

Dimension-wise Weighted Fusion. While ID embeddings effectively capture collaborative patterns for head items with abundant interactions, LLM embeddings provide more reliable semantic information for tail items with sparse data [38, 57]. To leverage these complementary strengths, we apply dimension-wise weighting to integrate the two embeddings, allowing the model to balance collaborative and semantic information at each dimension.

For the item v_i, we concatenate its ID embedding $\mathbf{e}_i^{ID}$ and LLM embedding $\mathbf{e}_i^{LLM}$ and feed them into the gating network:

$$\mathbf{g}_i = \sigma\left(\mathbf{W}_2(\mathbf{W}_1[\mathbf{e}_i^{ID} \oplus \mathbf{e}_i^{LLM}] + \mathbf{b}_1) + \mathbf{b}_2\right) \tag{3}$$

where $\mathbf{W}_1 \in \mathbb{R}^{2d\times 2d}$, $\mathbf{W}_2 \in \mathbb{R}^{d\times 2d}$, $\mathbf{b}_1 \in \mathbb{R}^{2d}$, and $\mathbf{b}_2 \in \mathbb{R}^d$ are the weight matrices and biases of the gating network. $\sigma(\cdot)$ denotes the Sigmoid function, and the output gating vector $\mathbf{g}_i \in \mathbb{R}^d$ takes values in $[0, 1]$ for each dimension. The final fused embedding is obtained via dimension-wise weighted interpolation:

$$\mathbf{f}_i = \mathbf{g}_i \odot \mathbf{e}_i^{ID} + (\mathbf{1} - \mathbf{g}_i) \odot \mathbf{e}_i^{LLM} \tag{4}$$

After applying this fusion to all items interacted by user u, we obtain the fused item embedding sequence $\mathbf{F}_u = [\mathbf{f}_1, \ldots, \mathbf{f}_{|s_u|}]$. The sequence is then fed into the SR encoder to produce the output representation of the sequence H_u.

4.2 Dual-Level Alignment

Existing LLM-based methods typically enhance ID embeddings directly leveraging LLM embeddings, without considering whether the structure of the two embedding spaces is aligned [21, 39]. This structural mismatch introduces conflicting signals when integrating collaborative and semantic information, undermining the effectiveness of LLM enhancement. To tackle this issue, we propose a dual-level alignment approach, which progressively aligns the two embedding spaces from point-to-point correspondences at the item level to structural consistency at the feature level.

Item-Level Alignment. The first level of alignment aims to establish point-to-point correspondences between ID and LLM embeddings of the same item across the two spaces. We introduce contrastive learning [4, 60] to achieve this. Specifically, for a training batch of K items, the item-level alignment loss is defined as:

$$\mathcal{L}_{ILA} = -\frac{1}{K}\sum_{i=1}^{K} \log \frac{\exp(\text{sim}(\mathbf{e}_i^{ID}, \mathbf{e}_i^{LLM})/\tau)}{\sum_{j=1}^{K} \exp(\text{sim}(\mathbf{e}_i^{ID}, \mathbf{e}_j^{LLM})/\tau)} \tag{5}$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity and τ is the temperature hyperparameter of contrastive learning. This loss function treats ID-LLM embedding pairs for the same item as positive samples, and pairs from different items as negative samples.

Feature-Level Alignment. Recent studies demonstrate that constraining the correlation structure of feature dimensions can effectively improve embedding quality [13, 65]. Inspired by this, we adopt the Barlow Twins [64] to achieve alignment at the feature-level. Specifically, we construct a cross-view correlation matrix that enforces high correlation between corresponding dimensions while decoupling different dimensions, achieving structural consistency between the ID and LLM embedding spaces.

Unlike item-level alignment, feature-level alignment requires calculating statistics at the batch dimension. Highly popular items,

due to their higher interaction frequency, dominate the normalization process, thereby weakening the alignment effect for tail items. Therefore, we use the median popularity p_{median} of batch items as a threshold to partition the batch item set into high-popularity and low-popularity item sets. The median strategy balanced group sizes and prevents the model training from being dominated by either group. The process is formalized as follows:

$$\mathcal{V}^{B_h} = \{i \mid p_i \geq p_{median}\}, \quad \mathcal{V}^{B_l} = \{i \mid p_i < p_{median}\} \tag{6}$$

where p_i denotes the popularity of item i. We illustrate the progress using $\mathcal{V}^{B_h}$ as an example, with $\mathcal{V}^{B_l}$ following the same process.

For the S item samples in the high-popularity group, we first standardize the ID embeddings and LLM embeddings along the batch dimension separately. Specifically, for each embedding matrix $\mathbf{E} \in \mathbb{R}^{S\times d}$, its standardized form is defined as follows:

$$\tilde{\mathbf{e}}_{n,i}^{B_h} = \frac{\mathbf{e}_{n,i}^{B_h} - \mu_i^{B_h}}{\sigma_i^{B_h}} \tag{7}$$

where $\mu_i^{B_h}$ and $\sigma_i^{B_h}$ respectively denote the mean and standard deviation of the i-th dimension of the embedding matrix $\mathbf{E}$. The normalized embedding matrices are denoted as $\tilde{\mathbf{E}}^{\text{ID}}, \tilde{\mathbf{E}}^{\text{LLM}} \in \mathbb{R}^{S\times d}$.

Based on the standardized embeddings, we conduct the cross-correlation matrix $\mathbf{C}^{B_h} \in \mathbb{R}^{d\times d}$, whose elements are defined as:

$$\mathbf{C}_{ij}^{B_h} = \frac{\sum_{s=1}^{S} \tilde{\mathbf{e}}_{s,i}^{ID} \tilde{\mathbf{e}}_{s,j}^{LLM}}{\sqrt{\sum_{s=1}^{S} (\tilde{\mathbf{e}}_{s,i}^{ID})^2} \sqrt{\sum_{s=1}^{S} (\tilde{\mathbf{e}}_{s,j}^{LLM})^2}} \tag{8}$$

This matrix reveals two types of dimensional correlations: diagonal elements $\mathbf{C}_{ii}^{B_h}$ measure the alignment of corresponding dimensions between ID and LLM embeddings, while off-diagonal elements $\mathbf{C}_{ij}^{B_h}$ $(i \neq j)$ measure redundancy across different dimensions.

To enhance correlation along corresponding dimensions while reducing redundancy across different dimensions, we define the alignment loss for the high-popularity item group as [64]:

$$\mathcal{L}_{FLA}^{B_h} = \underbrace{\sum_{i=1}^{d} (1 - \mathbf{C}_{ii}^{B_h})^2}_{\text{invariance}} + \lambda \underbrace{\sum_{i=1}^{d} \sum_{j\neq i} (\mathbf{C}_{ij}^{B_h})^2}_{\text{redundancy reduction}} \tag{9}$$

where the invariance term promotes alignment within the corresponding dimensions, the redundancy reduction term promotes decoupling between different dimensions, and λ controls the weight balance between the two terms. Similarly, we compute the feature-level alignment loss $\mathcal{L}_{FLA}^{B_l}$ for the low-popularity item group. The final feature-level alignment loss is the sum of the two losses:

$$\mathcal{L}_{FLA} = \mathcal{L}_{FLA}^{B_h} + \mathcal{L}_{FLA}^{B_l} \tag{10}$$

Curriculum Learning Scheduler. Effective representation learning typically follows a hierarchical principle from simple to complex [55], with item-level alignment and feature-level alignment precisely corresponding to this hierarchy: the former is a simple point-to-point task, while the latter is a complex alignment task over a space structure. If both constraints are imposed simultaneously during early training, premature space structure constraints may introduce noise, leading to suboptimal performance [16]. Therefore, inspired by curriculum learning [5, 17] we propose a cosine-scheduled to dynamically balance the two alignment objectives.

Table 1: The statistics of three datasets.

Dataset	# Users	# Items	Sparsity	Avg.length
Beauty	52,204	57,289	99.99%	7.56
Grocery	32,126	39,264	99.98%	8.57
Yelp	15,720	11,383	99.89%	12.23

Specifically, the weight of item-level alignment loss is defined as:

$$w_i(t) = (w_{\max} - w_{\min}) \cdot \frac{1 + \cos(2\pi t/T)}{2} + w_{\min} \tag{11}$$

where t denotes the training epoch, T is the oscillation period, and $w_{\max}$ and $w_{\min}$ control the upper and lower bounds of the weight, respectively. Through this oscillation, the two alignment mechanisms complement each other with dynamically adjusted emphasis, enabling progressive refinement of item embeddings.

Integrating the item-level alignment loss $\mathcal{L}_{ILA}$ and feature-level alignment loss $\mathcal{L}_{FLA}$, the overall alignment loss is:

$$\mathcal{L}_{align} = w_i(t) \cdot \mathcal{L}_{ILA} + (1 - w_i(t)) \cdot \mathcal{L}_{FLA} \tag{12}$$

4.3 Training and Inference

Training. We adopt the commonly used binary cross-entropy (BCE) loss to train the sequential recommendation models [12, 25, 63].

$$\begin{aligned} \mathcal{L}_{rec} &= \text{BCE}\left(H_u, E_u^+, E_u^-\right) \\ &= \left[\log\left(\sigma\left(H_u \cdot E_u^+\right)\right) + \log\left(1 - \sigma\left(H_u \cdot E_u^-\right)\right)\right] \end{aligned} \tag{13}$$

where H_u, E_u^+ and E_u^- denote the representations of the user, positive and negative items, respectively. $\sigma(\cdot)$ is the sigmoid function. To jointly optimize recommendation performance and embedding alignment , we combine the recommendation loss with the alignment loss. The overall training objective is:

$$\mathcal{L} = \mathcal{L}_{rec} + \alpha \cdot \mathcal{L}_{align} \tag{14}$$

where α is a hyperparameter that balances the two objectives.

Inference. During inference, the dual-level alignment has established structural consistency between the two embedding spaces and both embeddings are learned adequately. Therefore, to simplify the fusion process for efficiency, we use equal-weighted fusion of ID and LLM embeddings as final item embeddings for prediction.

4.4 Discussion

Traditional approaches [24, 26] leverage collaborative signals from head items to improve tail-item performance. However, for items with extremely sparse interactions, the insufficient contextual signals limit the effectiveness of such collaborative filtering-based knowledge transfer. Some works [2, 15] directly employ LLMs as recommenders, requiring LLM invocation at each inference step and introducing substantial computational overhead. Some works [22, 46, 47] use LLM embeddings as auxiliary signals to enhance ID embeddings during training but discard semantic information at inference time. Some works [21, 39] neglect structural consistency between embedding spaces when integrating the two embeddings. In contrast, our FAERec avoids high computational costs by using precomputed LLM embeddings. adaptive gating fusion leverages semantic knowledge to compensate for sparse collaborative signals in tail items. Dual-level alignment establishes structural consistency across embedding spaces, further improving the enhancement of

Table 2: Performance comparison of different methods on Backbone models. The best performance is bolded. The 'Impr.' represents the relative improvement of FAERec over the best baseline method.

Backbone		Beauty				Grocery				Yelp			
		Overall		Tail		Overall		Tail		Overall		Tail	
		H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
SASRec	Base	0.0331	0.0180	0.0113	0.0072	0.0332	0.0168	0.0086	0.0055	0.0240	0.0113	0.0060	0.0036
	CITIES	0.0305	0.0159	0.0106	0.0064	0.0286	0.0144	0.0074	0.0045	0.0240	0.0111	0.0049	0.0024
	LOAM	0.0300	0.0158	0.0115	0.0076	0.0285	0.0148	0.0089	0.0059	0.0122	0.0055	0.0060	0.0031
	MELT	0.0178	0.0088	0.0001	0.0000	0.0208	0.0101	0.0004	0.0003	0.0382	0.0188	0.0004	0.0002
	RLMRec-Con	0.0521	0.0287	0.0239	0.0132	0.0607	0.0344	0.0418	0.0248	0.0406	0.0212	0.0073	0.0041
	RLMRec-Gen	0.0439	0.0242	0.0122	0.0066	0.0471	0.0252	0.0119	0.0062	0.0391	0.0198	0.0051	0.0021
	LLMInit	0.0377	0.0200	0.0103	0.0064	0.0434	0.0225	0.0087	0.0049	0.0374	0.0198	0.0044	0.0018
	LLM-ESR	0.0269	0.0140	0.0049	0.0028	0.0298	0.0157	0.0051	0.0027	0.0412	0.0209	0.0040	0.0019
	FAERec	**0.0642**	**0.0372**	**0.0633**	**0.0368**	**0.0776**	**0.0446**	**0.0718**	**0.0421**	**0.0461**	**0.0225**	**0.0090**	**0.0047**
	Impr.	23.22%	29.62%	164.85%	178.79%	27.84%	29.65%	71.77%	69.76%	11.89%	6.13%	23.29%	14.63%
FMLP-Rec	Base	0.0277	0.0153	0.0066	0.0047	0.0285	0.0146	0.0053	0.0037	0.0153	0.0072	0.0029	0.0014
	CITIES	0.0238	0.0134	0.0062	0.0044	0.0252	0.0136	0.0048	0.0032	0.0153	0.0077	0.0015	0.0007
	LOAM	0.0225	0.0119	0.0057	0.0044	0.0221	0.0111	0.0048	0.0036	0.0115	0.0055	0.0022	0.0011
	MELT	0.0268	0.0143	0.0053	0.0030	0.0242	0.0122	0.0026	0.0014	0.0242	0.0119	0.0033	0.0013
	RLMRec-Con	0.0458	0.0262	0.0360	0.0217	0.0585	0.0322	0.0403	0.0229	0.0232	0.0114	0.0090	0.0048
	RLMRec-Gen	0.0533	0.0305	0.0269	0.0151	0.0550	0.0304	0.0285	0.0166	0.0178	0.0087	0.0026	0.0014
	LLMInit	0.0383	0.0205	0.0096	0.0061	0.0406	0.0208	0.0091	0.0053	0.0265	0.0130	0.0038	0.0019
	LLM-ESR	0.0224	0.0118	0.0051	0.0031	0.0274	0.0144	0.0081	0.0046	**0.0308**	**0.0159**	0.0040	0.0017
	FAERec	**0.0622**	**0.0361**	**0.0645**	**0.0379**	**0.0749**	**0.0425**	**0.0737**	**0.0426**	0.0292	0.0149	**0.0097**	**0.0051**
	Impr.	16.70%	18.36%	79.17%	74.65%	28.03%	31.99%	82.88%	86.03%	-5.19%	-6.29%	7.78%	6.25%
LRURec	Base	0.0349	0.0192	0.0108	0.0069	0.0351	0.0184	0.0080	0.0046	0.0288	0.0144	0.0033	0.0015
	CITIES	0.0330	0.0175	0.0101	0.0062	0.0335	0.0175	0.0077	0.0047	0.0247	0.0115	0.0031	0.0013
	LOAM	0.0302	0.0160	0.0110	0.0073	0.0296	0.0150	0.0081	0.0050	0.0248	0.0119	0.0030	0.0015
	MELT	0.0242	0.0120	0.0018	0.0008	0.0251	0.0125	0.0008	0.0004	0.0387	0.0187	0.0004	0.0002
	RLMRec-Con	0.0372	0.0205	0.0117	0.0073	0.0602	0.0335	0.0239	0.0135	0.0344	0.0172	0.0035	0.0020
	RLMRec-Gen	0.0418	0.0229	0.0098	0.0058	0.0458	0.0249	0.0070	0.0041	0.0394	0.0202	0.0031	0.0013
	LLMInit	0.0383	0.0205	0.0096	0.0061	0.0451	0.0234	0.0089	0.0051	0.0415	0.0211	0.0057	0.0026
	LLM-ESR	0.0277	0.0144	0.0045	0.0024	0.0333	0.0173	0.0054	0.0030	0.0384	0.0194	0.0029	0.0013
	FAERec	**0.0644**	**0.0374**	**0.0617**	**0.0357**	**0.0759**	**0.0434**	**0.0675**	**0.0382**	**0.0433**	**0.0219**	**0.0086**	**0.0045**
	Impr.	54.07%	63.32%	427.35%	389.04%	26.08%	29.55%	182.43%	182.96%	4.34%	3.79%	50.88%	73.08%

LLM embeddings for recommendation. Moreover, our approach is model-agnostic and can be adapted to various SR models.

5 Experiments

5.1 Experiment Setup

5.1.1 Datasets. The experiments are conducted on three datasets. Beauty and Grocery datasets are sourced from Amazon [44] and correspond to the "Beauty" and "Grocery and Gourmet Food" categories, respectively. Yelp [2] contains user check-in histories along with their reviews. The maximum sequence length is set to 50 for all datasets. Table 1 presents the detailed statistics. We follow the preprocessing of the previous SR works [6, 25, 26, 39]. Since the sequential recommendation is commonly used with implicit feedback, we treat all rating records or reviews as interactions. Then, we exclude accounts with fewer than three interaction records, as the cold-start user problem is beyond our research scope.

[2]https://www.yelp.com/dataset

5.1.2 Backbones. To verify the generalizability of our FAERec, we selected three representative general SR models as our backbones.

- SASRec [25] leverages a unidirectional multi-head self-attention layers mechanism to model the casual transition correlations for the next item in the behavior sequence.
- FMLP-Rec [72] introduces an all-MLP model with learnable filters for noise reduction in sequential recommendation.
- LRURec [63] proposes a linear recurrence model based on matrix diagonalization to capture transition patterns while improving training efficiency through parallel training.

5.1.3 Baselines. We compare FAERec against various baseline methods, including three traditional long-tail models and three recent LLM-based methods.

- CITIES [24] trains an embedding inference function using head item contexts to improve tail item quality.
- MELT [26] designs a mutually reinforcing dual-branch framework to tackle the long-tail user and item challenge. One branch

Table 3: The ablation study results on the Beauty dataset with three sequential recommendation backbones.

Variant	SASRec				FMLP-Rec				LRURec			
	Overall		Tail		Overall		Tail		Overall		Tail	
	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
FAERec	**0.0642**	**0.0372**	**0.0633**	**0.0368**	0.0622	**0.0361**	**0.0645**	**0.0379**	**0.0644**	**0.0374**	**0.0617**	**0.0357**
w/o AGF	0.0534	0.0267	0.0359	0.0182	**0.0624**	0.0355	0.0516	0.0291	0.0561	0.0289	0.0369	0.0185
w/o ILA	0.0568	0.0318	0.0473	0.0260	0.0585	0.0343	0.0567	0.0331	0.0519	0.0289	0.0353	0.0192
w/o FLA	0.0626	0.0365	0.0603	0.0355	0.0597	0.0353	0.0630	0.0376	0.0632	0.0369	0.0595	0.0346
w/o CLS	0.0554	0.0309	0.0436	0.0239	0.0576	0.0330	0.0538	0.0305	0.0529	0.0292	0.0401	0.0218
w/o PG	0.0636	0.0367	0.0601	0.0350	0.0618	0.0357	0.0642	0.0373	0.0636	0.0369	0.0583	0.0339
Base	0.0331	0.0180	0.0113	0.0072	0.0277	0.0153	0.0066	0.0047	0.0349	0.0192	0.0108	0.0069

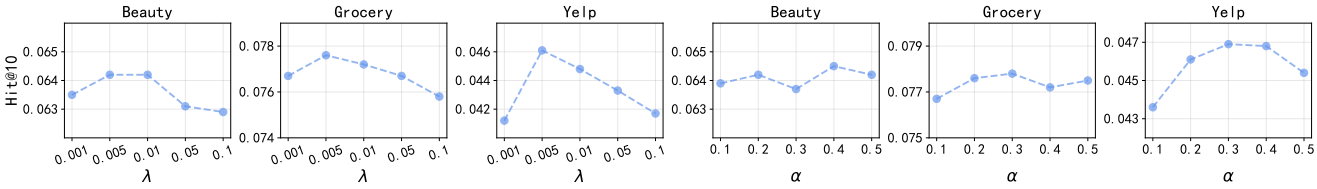


Figure 3: Sensitivity analysis of hyperparameter λ and α on three dataset with SASRec as the backbone.

transfers knowledge from head users to tail users, while the other transfers patterns from head items to tail items.

- LOAM [61] employs NicheWalk augmentation to capture session-level patterns with global context and Tail Session Mixup to synthesize interactions, enriching representations for tail items.
- RLMRec [47], originally designed for collaborative filtering, is adapted to sequential recommendation in our work.It employs two knowledge transfer strategies to maximize mutual information between embeddings. RLMRec-Con maps LLM embeddings to collaborative space, while RLMRec-Gen maps randomly masked ID embeddings to semantic space for reconstruction, both using similarity-based regularization.
- LLMInit [18, 22, 46] represents a category of methods that use language embeddings to initialize ID embeddings. Representative works include LLM2Bert4Rec [18], SAID [22], and Elephant [46], which initialize the item embedding layer with LLM embeddings and subsequently fine-tune them using interaction data.
- LLM-ESR [39] combines semantic knowledge from LLMs with collaborative signals for tail items through dual-view modeling, and introduce retrieval-augmented distillation for tail users.

5.1.4 Evaluation Settings. We adopt the leave-one-out strategy to partition sequences into training, validation, and test sets. Following previous practice [8, 9, 30, 60, 63], we rank the prediction over the whole item set rather than negative sampling. The evaluation metrics include Hit Ratio@K (H@K) and Normalized Discounted Cumulative Gain@K (N@K). We report results with K $\in \{5, 10, 20\}$.

5.1.5 Implementation Details. For all baselines, we adopt the implementation provided by the authors. For training, we set the batch size to 256 across all datasets and use the Adam [28] optimizer with a learning rate of 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Following previous works [21, 39], we set the embedding size to 128. To ensure fair comparisons, we carefully set and tune all other hyperparameters of each method as reported and suggested in the original papers. For FAERec, we tune the λ, α in the range of $\{0.001, 0.005, 0.01, 0.05, 0.1\}$, $\{0.1, 0.2, 0.3, 0.4, 0.5\}$, respectively. Furthermore, the embeddings of LLMs are derived from the API named "text-embedding-ada-002" provided by OpenAI.

5.2 Overall Comparison

The experimental results of our FAERec and different baseline methods with various backbone SR models are presented in Table 2 [3].

We have the following observations and conclusions:

- Among the backbone models, LRURec achieves the best performance owing to its efficient linear recurrent modeling capability. For traditional long-tail methods, although they improve performance on tail items, their overall performance sometimes falls short of the base models due to the seesaw problem [39] (i.e., overemphasizing tail items leads to degraded performance on head items). Notably, LLM-based methods consistently outperform traditional long-tail approaches, demonstrating the advantage of semantic information in handling extremely sparse data and the limitations of relying solely on collaborative signals.
- Among LLM-based methods, RLMRec-Con and RLMRec-Gen achieve leading overall performance in most cases, validating the effectiveness of cross-view knowledge transfer. RLMRec-Gen underperforms RLMRec-Con on tail items, as its masked reconstruction strategy further aggravates sparsity for tail items. In contrast, LLMInit and LLM-ESR perform relatively poorly. The former only initializes with semantic information and discards it during training, while the latter overlooks consistency of embedding spaces when combining ID and LLM embeddings.
- For our proposed FAERec, it achieves the best performance in nearly all cases. Notably, for tail items, the model achieves optimal performance across all cases. Through the adaptive fusion and

[3] For MELT [26], despite strictly following the authors' provided code and replicating the parameter settings and tuning ranges reported in the original paper, its performance consistently exhibits anomalies and underperforms in certain scenarios. We have observed similar phenomena in other published papers [33, 34, 39].

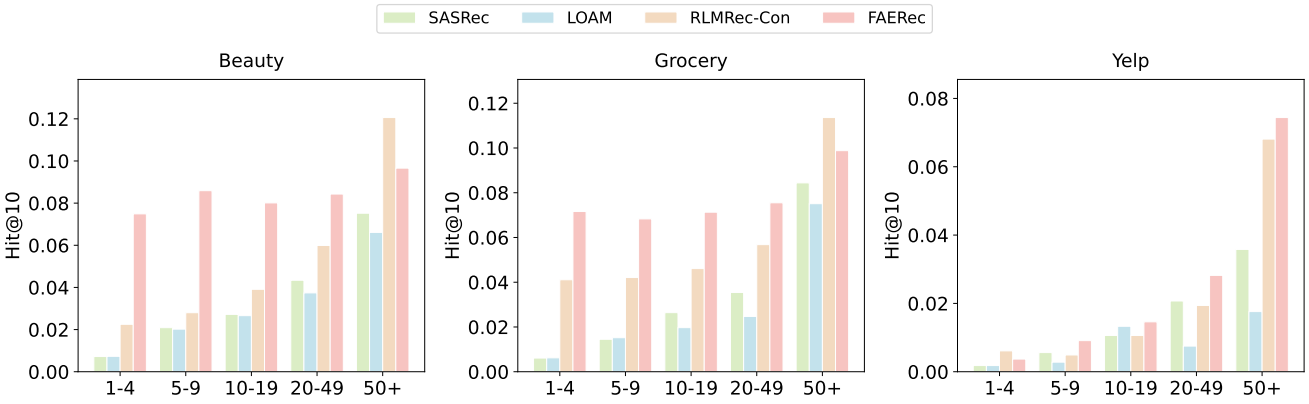


Figure 4: The group analysis results on three dataset with SASRec as the backbone.

Table 4: Results on more sequential recommendation backbones.

Backbone	Beauty				Grocery				Yelp			
	Overall		Tail		Overall		Tail		Overall		Tail	
	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
NARM	0.0197	0.0097	0.0017	0.0010	0.0219	0.0114	0.0016	0.0008	0.0387	0.0195	0.0015	0.0009
w/FAERec	**0.0429**	**0.0226**	**0.0331**	**0.0176**	**0.0540**	**0.0286**	**0.0400**	**0.0213**	**0.0456**	**0.0223**	**0.0051**	**0.0021**
LightSANs	0.0252	0.0143	0.0059	0.0046	0.0291	0.0150	0.0052	0.0038	0.0167	0.0080	0.0024	0.0016
w/FAERec	**0.0597**	**0.0353**	**0.0614**	**0.0364**	**0.0714**	**0.0406**	**0.0727**	**0.0425**	**0.0254**	**0.0125**	**0.0115**	**0.0057**
gMLP	0.0271	0.0163	0.0076	0.0058	0.0299	0.0157	0.0058	0.0040	0.0203	0.0093	0.0013	0.0007
w/FAERec	**0.0571**	**0.0325**	**0.0547**	**0.0311**	**0.0694**	**0.0383**	**0.0622**	**0.0340**	**0.0337**	**0.0169**	**0.0099**	**0.0048**

dual-level alignment framework, FAERec effectively integrates collaborative signals and semantic information while maintaining consistency between embedding spaces. Even for scenarios with sparse interactions, FAERec leverages semantic information to compensate for the insufficiency of collaborative signals, thereby significantly enhancing the performance of tail items.

5.3 Ablation Study

We conducted an ablation study to validate the effectiveness of various components within FAERec. We compared FAERec with the following variants: 1) w/o AGF: Removing adaptive gating fusion. 2) w/o ILA: Removing item-level alignment. 3) w/o FLA: Removing feature-level alignment. 4) w/o CLS: Setting both alignment losses to fixed weights without curriculum learning scheduler. 5) w/o PG: Removing popularity-based grouping in feature-level alignment.

As shown in Table 3, all modules contribute to performance improvements. Removing adaptive gating fusion leads to significant performance degradation, particularly on tail items, highlighting the importance of incorporating semantic information into item embeddings for tail items. After removing either item-level or feature-level alignment, the models' performance metrics declined, demonstrating that both alignment strategies jointly facilitate model training. Notably, removing item-level alignment severely degrades tail-item performance, underscoring the need to establish basic item correspondences when combining heterogeneous embeddings. Performance decreases after removing the curriculum learning scheduler, even falling below that w/o FLA, verifying the importance of proper scheduling between the two alignments. If feature-level alignment is introduced prematurely, it injects noise during training and may even disrupt the correspondences established by item-level alignment. Additionally, popularity-based grouping prevents high-popularity items from dominating the alignment of low-popularity items at the feature level. Overall, these results validate that each component effectively addresses specific challenges for tail-item.

5.4 Hyperparameter Investigation

We further investigated the redundancy reduction weight λ for feature-level alignment and the total alignment loss weight α. The experimental results are shown in Figure 3. As λ increased from 0.001 to 0.1, the overall Hit@10 value of FAERec first rose and then declined, with the optimal value occurring in the range $\lambda \in \{0.005, 0.01\}$. This indicates that moderate redundancy reduction constraints can effectively maintain high alignment within corresponding dimensions while reducing information redundancy between different dimensional, thereby promoting consistency in the dimensional structures of the two embedding spaces.

For the alignment loss weight α, we can observe that its optimal value typically lies within the range $\{0.3, 0.4\}$. When α is within this range, the alignment loss effectively establishes a stable and consistent corresponding between the ID and LLM embedding space, thereby promoting the effective integration of collaborative signals and semantic information. However, when α is too small, the alignment constraint fails to establish structural consistency between the two embedding spaces sufficiently. Conversely, when α is excessively large, overly stringent alignment can lead to convergence difficulties, causing the model to overemphasize space structure

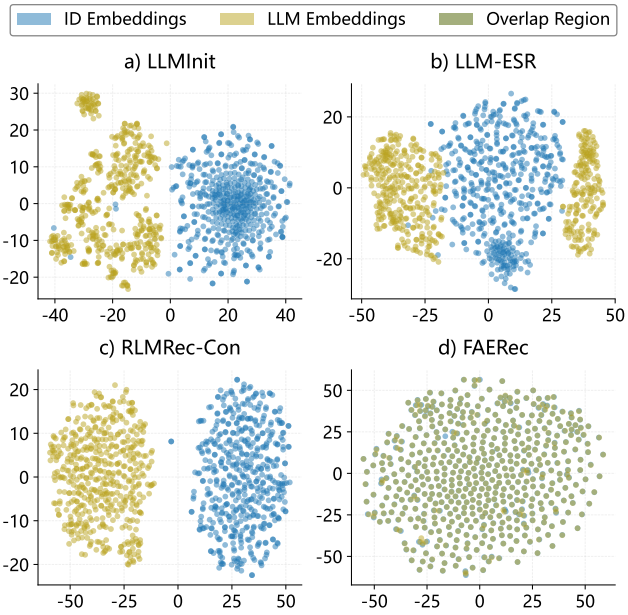


Figure 5: Visualization of ID and LLM embedding spaces on the Beauty dataset with SASRec as the backbone.

alignment at the expense of primary recommendation objectives, thereby degrading overall performance.These observations guide our hyperparameter configuration in practice.

5.5 Group Analysis

To investigate the long-tail problem more comprehensively, we divided items by their popularity into 5 groups. As shown in Figure 4, we observe that FAERec can improve performance over the backbone across items of any popularity. In the 1-4 group, LOAM's performance on long-tail item groups lags significantly behind LLM-based models, demonstrating the limitations of traditional long-tail methods for extreme cold-start items. Compared with RLMRec-Con, our FAERec exhibits more performance improvements on long-tail items, especially for those in group 1-4 and group 5-9. However, FAERec slightly underperforms RLMRec-Con for popular items (e.g., 50+ group), indicating the inherent challenge of balancing performance between popular and long-tail items.

5.6 Visualization

To verify the proposed structural inconsistency between the ID and LLM embedding spaces and the alignment effectiveness of FAERec, we visualize the embeddings of the same sampled item set using t-SNE in Figure 5. We observe that LLMInit exhibits a scattered and uneven distribution since its LLM embeddings remain untrained. LLM-ESR lacks alignment strategies, resulting in significant discrepancies between ID and LLM embeddings. By introducing cross-view contrastive learning, RLMRec-Con achieves similar distribution patterns across ID and LLM embeddings. In contrast, FAERec goes further: through the dual-level alignment framework, it not only produces more uniform distributions but also establishes high structural consistency across the two embedding spaces, thereby improving the quality of item embeddings.

Table 5: Diversity comparison of different methods on the Beauty dataset with three SR backbones.

Method	SASRec		FMLP-Rec		LRURec	
	Cov@10	TCov@10	Cov@10	TCov@10	Cov@10	TCov@10
Base	0.6972	0.6266	0.5273	0.4267	0.6679	0.5920
LOAM	0.6419	0.5634	0.5092	0.4057	0.6768	0.6036
RLMRec-Con	0.5019	0.4141	0.7010	0.6442	0.5477	0.4554
LLM-ESR	0.3513	0.2251	0.5176	0.4215	0.4075	0.2831
FAERec	**0.8538**	**0.8249**	**0.8571**	**0.8298**	**0.8612**	**0.8315**

5.7 Generalizing to More Backbones

To further verify the generalization capability of FAERec, we apply it to more SR backbones with different architectures:

- NARM [31]: An RNN-based model that integrates attention mechanisms into recurrent neural networks to capture both global sequential behaviors and local key intentions.
- LightSANs [14]: A lightweight self-attention model that projects user interests through low-rank factorized self-attention and generates context-aware item representations.
- gMLP [35]: An MLP-based model that replaces self-attention with spatial gating units, capturing sequential dependencies through cross-position multi-layer perceptrons.

As shown in Table 4, we observe that FAERec achieves significant improvements across all backbone models consistently. For overall performance, FAERec achieves average improvements of 124.1%, 148.3%, and 48.0% across the three datasets, respectively. For long-tail item recommendation, the improvements are even more substantial, reaching 1032.5%, 1500.2%, and 376.0%, respectively. These results validate the effectiveness and generalization capability of our method, particularly demonstrating outstanding performance in long-tail item recommendation scenarios.

5.8 Diversity Analysis

Diverse recommendations deliver surprise and discovery value, enhancing long-term user satisfaction [27]. Therefore, we further evaluate the diversity of FAERec's recommendation. Following previous research [42, 61], we adopt Coverage@K (Cov@K) and Tail_Coverage@K (TCov@K), which measure the number of distinct items appearing in the top-K recommendations and the number of long-tail items, respectively. Their definitions are as follows:

$$\begin{aligned} \text{Coverage@K} &= \frac{|\bigcup_{u\in\mathcal{U}} L_K(u)|}{|\mathcal{V}|} \\ \text{Tail_Coverage@K} &= \frac{|\bigcup_{u\in\mathcal{U}} L_K^{\text{tail}}(u)|}{|\mathcal{V}_{\text{tail}}|} \end{aligned} \tag{15}$$

where $L_K(u)$ denotes the top-K recommendation list generated for user u. Specifically, $L_K^{\text{tail}}(u)$ represents the subset of $L_K(u)$ that contains items belonging to the long-tail item set $\mathcal{V}_{\text{tail}}$.

As shown in Table 5, although LOAM and RLMRec-Con improve the prediction accuracy for long-tail items, their contribution to recommendation diversity remains limited. In contrast, our proposed FAERec not only achieves improvements in tail-item performance but also significantly enhances recommendation diversity by effectively integrating collaborative signals with semantic information to compensate for the inherent disadvantages of long-tail items.

6 CONCLUSION

This paper introduces FAERec, a fusion and alignment enhancement framework with LLMs tailored for tail item sequential recommendations. To tackle the challenges of long-tail items, FAERec improves item embedding quality and enhances long-tail recommendation performance by effectively integrating collaborative signals with semantic information. Specifically, we design an adaptive gated fusion mechanism to capture fine-grained complementary patterns between the two embeddings. To further enhance fusion effectiveness, we introduce a dual-layer alignment framework, item-level alignment and feature-level alignment, establishing point-to-point correspondences and spatial structural consistency between the two embedding spaces. Additionally, a curriculum learning scheduler balances the two alignment objectives to ensure training stability. We conducted experiments on three real-world datasets to comprehensively validate the effectiveness and generalizability of FAERec.

7 Acknowledgments

This research is partially supported by the National Natural Science Foundation of China under Grant No. 62576083. This research is also supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 1 grant, and funded through the SMU-SUTD Internal Research Grant Call (SMU-SUTD 2023_02_01), and in part by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (Award No. MOE-T2EP201230015). The authors greatly appreciate the anonymous reviewers for their valuable comments.

References

[1] Ido Amos, Jonathan Berant, and Ankit Gupta. 2023. Never train from scratch: Fair comparison of long-sequence models requires data-driven priors. *arXiv preprint arXiv:2310.02980* (2023).

[2] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. 2023. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM conference on recommender systems*. 1007–1014.

[3] George EP Box and R Daniel Meyer. 1986. An analysis for unreplicated fractional factorials. *Technometrics* 28, 1 (1986), 11–18.

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PmLR, 1597–1607.

[5] Yudong Chen, Xin Wang, Miao Fan, Jizhou Huang, Shengwen Yang, and Wenwu Zhu. 2021. Curriculum meta-learning for next POI recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2692–2702.

[6] Yizhou Dang, Yuting Liu, Enneng Yang, Minhan Huang, Guibing Guo, Jianzhe Zhao, and Xingwei Wang. 2025. Data augmentation as free lunch: Exploring the test-time augmentation for sequential recommendation. In *SIGIR*. 1466–1475.

[7] Yizhou Dang, Zhifu Wei, Minhan Huang, Lianbo Ma, Jianzhe Zhao, Guibing Guo, and Xingwei Wang. 2026. Tail-Aware Data Augmentation for Long-Tail Sequential Recommendation. *arXiv preprint arXiv:2601.10933* (2026).

[8] Yizhou Dang, Enneng Yang, Guibing Guo, Linying Jiang, Xingwei Wang, Xiaoxiao Xu, Qinghui Sun, and Hong Liu. 2023. TiCoSeRec: Augmenting data to uniform sequences by time intervals for effective recommendation. *TKDE* (2023).

[9] Yizhou Dang, Enneng Yang, Guibing Guo, Linying Jiang, Xingwei Wang, Xiaoxiao Xu, Qinghui Sun, and Hong Liu. 2023. Uniform sequence better: Time interval aware data augmentation for sequential recommendation. In *AAAI*.

[10] Yizhou Dang, Enneng Yang, Yuting Liu, Guibing Guo, Linying Jiang, Jianzhe Zhao, and Xingwei Wang. 2024. Data Augmentation for Sequential Recommendation: A Survey. *arXiv preprint arXiv:2409.13545* (2024).

[11] Yizhou Dang, Enneng Yang, Yuting Liu, Jianzhe Zhao, Xingwei Wang, and Guibing Guo. 2026. Exploring and Tailoring the Test-Time Augmentation for Sequential Recommendation. *TPAMI* (2026).

[12] Yizhou Dang, Jiahui Zhang, Yuting Liu, Enneng Yang, Yuliang Liang, Guibing Guo, Jianzhe Zhao, and Xingwei Wang. 2024. Augmenting Sequential Recommendation with Balanced Relevance and Diversity. *arXiv preprint arXiv:2412.08300* (2024).

[13] Aleksandr Ermolov, Aliaksandr Siarohin, Enver Sangineto, and Nicu Sebe. 2021. Whitening for self-supervised representation learning. In *International conference on machine learning*. PMLR, 3015–3024.

[14] Xinyan Fan, Zheng Liu, Jianxun Lian, Wayne Xin Zhao, Xing Xie, and Ji-Rong Wen. 2021. Lighter and better: low-rank decomposed self-attention networks for next-item recommendation. In *SIGIR*. 1733–1737.

[15] Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *arXiv preprint arXiv:2303.14524* (2023).

[16] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. 2017. Accurate, large minibatch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677* (2017).

[17] Alex Graves, Marc G Bellemare, Jacob Menick, Remi Munos, and Koray Kavukcuoglu. 2017. Automated curriculum learning for neural networks. In *international conference on machine learning*. Pmlr, 1311–1320.

[18] Jesse Harte, Wouter Zorgdrager, Panos Louridas, Asterios Katsifodimos, Dietmar Jannach, and Marios Fragkoulis. 2023. Leveraging large language models for sequential recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 1096–1102.

[19] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *ICDM*. IEEE, 191–200.

[20] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. *arXiv preprint arXiv:1511.06939* (2015).

[21] Guoqing Hu, An Zhang, Shuo Liu, Zhibo Cai, Xun Yang, and Xiang Wang. 2025. Alphafuse: Learn id embeddings for sequential recommendation in null space of language embeddings. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1614–1623.

[22] Jun Hu, Wenwen Xia, Xiaolu Zhang, Chilin Fu, Weichang Wu, Zhaoxin Huan, Ang Li, Zuoli Tang, and Jun Zhou. 2024. Enhancing sequential recommendation via llm-based semantic embedding learning. In *Companion Proceedings of the ACM Web Conference 2024*. 103–111.

[23] Yidan Hu, Yong Liu, Chunyan Miao, and Yuan Miao. 2022. Memory bank augmented long-tail sequential recommendation. In *CIKM*. 791–801.

[24] Seongwon Jang, Hoyeop Lee, Hyunsouk Cho, and Sehee Chung. 2020. Cities: Contextual inference of tail-item embeddings for sequential recommendation. In *ICMD*. IEEE, 202–211.

[25] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *ICDM*. IEEE, 197–206.

[26] Kibum Kim, Dongmin Hyun, Sukwon Yun, and Chanyoung Park. 2023. Melt: Mutual enhancement of long-tailed user and item for sequential recommendation. In *Proceedings of the 46th international ACM SIGIR conference on Research and development in information retrieval*. 68–77.

[27] Yejin Kim, Kwangseob Kim, Chanyoung Park, and Hwanjo Yu. 2019. Sequential and Diverse Recommendation with Long Tail.. In *IJCAI*, Vol. 19. 2740–2746.

[28] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[29] Anastasiia Klimashevskaia, Dietmar Jannach, Mehdi Elahi, and Christoph Trattner. 2024. A survey on popularity bias in recommender systems. *User Modeling and User-Adapted Interaction* 34, 5 (2024), 1777–1834.

[30] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *KDD*. 1748–1757.

[31] Jing Li, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Tao Lian, and Jun Ma. 2017. Neural attentive session-based recommendation. In *CIKM*. 1419–1428.

[32] Yaoyiran Li, Xiang Zhai, Moustafa Alzantot, Keyi Yu, Ivan Vulić, Anna Korhonen, and Mohamed Hammad. 2024. Calrec: Contrastive alignment of generative llms for sequential recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 422–432.

[33] Zihao Li, Yakun Chen, Tong Zhang, and Xianzhi Wang. 2025. Reembedding and Reweighting are Needed for Tail Item Sequential Recommendation. In *Proceedings of the ACM on Web Conference 2025*. 4925–4936.

[34] Zihao Li, Chao Yang, Tong Zhang, Yakun Chen, Xianzhi Wang, Guandong Xu, and Daoyi Dong. 2025. Listwise Preference Alignment Optimization for Tail Item Recommendation. *arXiv preprint arXiv:2507.02255* (2025).

[35] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. 2021. Pay attention to mlps. *NIPS* 34 (2021), 9204–9215.

[36] Junkang Liu, Yuanyuan Liu, Fanhua Shang, Hongying Liu, Jin Liu, and Wei Feng. 2025. Improving Generalization in Federated Learning with Highly Heterogeneous Data via Momentum-Based Stochastic Controlled Weight Averaging. In *Forty-second International Conference on Machine Learning*.

[37] Junkang Liu, Fanhua Shang, Kewen Zhu, Hongying Liu, Yuanyuan Liu, and Jin Liu. 2025. FedAdamW: A Communication-Efficient Optimizer with Convergence and Generalization Guarantees for Federated Large Models. *arXiv preprint arXiv:2510.27486* (2025).

[38] Qidong Liu, Xian Wu, Wanyu Wang, Yejing Wang, Yuanshao Zhu, Xiangyu Zhao, Feng Tian, and Yefeng Zheng. 2025. Llmemb: Large language model can be a good embedding generator for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 12183–12191.

[39] Qidong Liu, Xian Wu, Yejing Wang, Zijian Zhang, Feng Tian, Yefeng Zheng, and Xiangyu Zhao. 2024. Llm-esr: Large language models enhancement for long-tailed sequential recommendation. *Advances in Neural Information Processing Systems* 37 (2024), 26701–26727.

[40] Qidong Liu, Xiangyu Zhao, Yuhao Wang, Yejing Wang, Zijian Zhang, Yuqi Sun, Xiang Li, Maolin Wang, Pengyue Jia, Chong Chen, et al. 2025. Large Language Model Enhanced Recommender Systems: Methods, Applications and Trends. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*. 6096–6106.

[41] Qidong Liu, Xiangyu Zhao, Yejing Wang, Zijian Zhang, Howard Zhong, Chong Chen, Xiang Li, Wei Huang, and Feng Tian. 2025. Bridge the domains: Large language models enhanced cross-domain sequential recommendation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1582–1592.

[42] Siyi Liu and Yujia Zheng. 2020. Long-tail session-based recommendation. In *Proceedings of the 14th ACM conference on recommender systems*. 509–514.

[43] Yifan Liu, Kangning Zhang, Xiangyuan Ren, Yanhua Huang, Jiarui Jin, Yingjie Qin, Ruilong Su, Ruiwen Xu, Yong Yu, and Weinan Zhang. 2024. Alignrec: Aligning and training in multimodal recommendations. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 1503–1512.

[44] Julian McAuley, Rahul Pandey, and Jure Leskovec. 2015. Inferring networks of substitutable and complementary products. In *KDD*. 785–794.

[45] Shilin Qu, Fajie Yuan, Guibing Guo, Liguang Zhang, and Wei Wei. 2022. CmnRec: Sequential Recommendations with Chunk-accelerated Memory Network. *TKDE* (2022).

[46] Zekai Qu, Ruobing Xie, Chaojun Xiao, Zhanhui Kang, and Xingwu Sun. 2024. The elephant in the room: rethinking the usage of pre-trained language model in sequential recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems*. 53–62.

[47] Xubin Ren, Wei Wei, Lianghao Xia, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. 2024. Representation learning with large language models for recommendation. In *Proceedings of the ACM web conference 2024*. 3464–3475.

[48] 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*. 1441–1450.

[49] Yatong Sun, Bin Wang, Zhu Sun, and Xiaochun Yang. 2021. Does Every Data Instance Matter? Enhancing Sequential Recommendation by Eliminating Unreliable Data.. In *IJCAI*. 1579–1585.

[50] Yatong Sun, Bin Wang, Zhu Sun, Xiaochun Yang, and Yan Wang. 2023. Theoretically guaranteed bidirectional data rectification for robust sequential recommendation. *Advances in Neural Information Processing Systems* 36 (2023), 2850–2876.

[51] Yatong Sun, Xiaochun Yang, Zhu Sun, Yan Wang, Bin Wang, and Xinghua Qu. 2025. LLM4RSR: Large Language Models as Data Correctors for Robust Sequential Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 12604–12612.

[52] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *WSDM*. 565–573.

[53] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).

[54] A Vaswani. 2017. Attention is all you need. *NIPS* (2017).

[55] Xin Wang, Yudong Chen, and Wenwu Zhu. 2021. A survey on curriculum learning. *IEEE transactions on pattern analysis and machine intelligence* 44, 9 (2021), 4555–4576.

[56] Yuhao Wang, Junwei Pan, Pengyue Jia, Wanyu Wang, Maolin Wang, Zhixiang Feng, Xiaotian Li, Jie Jiang, and Xiangyu Zhao. 2025. Pre-train, Align, and Disentangle: Empowering Sequential Recommendation with Large Language Models. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1455–1465.

[57] Yuhao Wang, Junwei Pan, Xinhang Li, Maolin Wang, Yuan Wang, Yue Liu, Dapeng Liu, Jie Jiang, and Xiangyu Zhao. 2025. Empowering large language model for sequential recommendation via multimodal embeddings and semantic ids. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 3209–3219.

[58] Yuhao Wang, Yichao Wang, Zichuan Fu, Xiangyang Li, Wanyu Wang, Yuyang Ye, Xiangyu Zhao, Huifeng Guo, and Ruiming Tang. 2024. Llm4msr: An llm-enhanced paradigm for multi-scenario recommendation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 2472–2481.

[59] Likang Wu, Zhaopeng Qiu, Zhi Zheng, Hengshu Zhu, and Enhong Chen. 2024. Exploring large language model for graph data understanding in online job recommendations. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 9178–9186.

[60] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *ICDE*. IEEE, 1259–1273.

[61] Heeyoon Yang, YunSeok Choi, Gahyung Kim, and Jee-Hyong Lee. 2023. LOAM: improving long-tail session-based recommendation via niche walk augmentation and tail session mixup. In *Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval*. 527–536.

[62] Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M Jose, and Xiangnan He. 2019. A simple convolutional generative network for next item recommendation. In *WSDM*. 582–590.

[63] Zhenrui Yue, Yueqi Wang, Zhankui He, Huimin Zeng, Julian McAuley, and Dong Wang. 2024. Linear recurrent units for sequential recommendation. In *Proceedings of the 17th ACM international conference on web search and data mining*. 930–938.

[64] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. 2021. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*. PMLR, 12310–12320.

[65] Dan Zhang, Yangliao Geng, Wenwen Gong, Zhongang Qi, Zhiyu Chen, Xing Tang, Ying Shan, Yuxiao Dong, and Jie Tang. 2024. Recdcl: Dual contrastive learning for recommendation. In *Proceedings of the ACM Web Conference 2024*. 3655–3666.

[66] Fan Zhang, Gongguan Chen, Hua Wang, and Caiming Zhang. 2024. CF-DAN: Facial-expression recognition based on cross-fusion dual-attention network. *Computational Visual Media* 10, 3 (2024), 593–608.

[67] Chuang Zhao, Xinyu Li, Ming He, Hongke Zhao, and Jianping Fan. 2023. Sequential Recommendation via an Adaptive Cross-domain Knowledge Decomposition. In *CIKM*. 3453–3463.

[68] Chuang Zhao, Hui Tang, Hongke Zhao, Xiaofang Zhou, and Xiaomeng Li. 2025. Grounded by Experience: Generative Healthcare Prediction Augmented with Hierarchical Agentic Retrieval. *arXiv preprint arXiv:2511.13293* (2025).

[69] Chuang Zhao, Hongke Zhao, Ming He, Jian Zhang, and Jianping Fan. 2023. Cross-domain recommendation via user interest alignment. In *WWW*. 887–896.

[70] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* 1, 2 (2023).

[71] Xiaoquan Zhi, Hongke Zhao, Likang Wu, Chuang Zhao, and Hengshu Zhu. 2025. Reinventing Clinical Dialogue: Agentic Paradigms for LLM Enabled Healthcare Communication. *arXiv preprint arXiv:2512.01453* (2025).

[72] Kun Zhou, Hui Yu, Wayne Xin Zhao, and Ji-Rong Wen. 2022. Filter-enhanced MLP is all you need for sequential recommendation. In *WWW*. 2388–2399.