ProRec-Video: Guiding Hierarchical Interest Transitions for Proactive Short Video Recommendation with Dynamic Feedback Adaptation

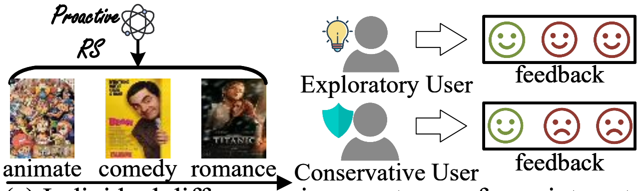
Weizhi Chen[1], **Baoyun Peng**[2], **Bo Liu**[1, 2], **Xingkong Ma**[1]*, **Houjie Qiu**[1]
[1]College of Computer Science and Technology, National University of Defense Technology, Changsha, Hunan, China
[2]Academy of Military Science, Beijing, China
{chenweizhi17, kyle.liu, maxingkong, qiuhoujie}@nudt.edu.cn, pengbaoyun13@alumni.nudt.edu.cn

Abstract

Traditional short video recommendations primarily enhance user stickiness by reinforcing existing user preferences, potentially leading to information cocoons. Conversely, proactive recommendations aim to diversify user interests by exposing users to content beyond their historical preferences. However, current proactive approaches face three limitations: (1) homogeneous receptivity assumption, neglecting individual differences in users' openness to new interests; (2) short-term item exposure without interest anchoring, focusing on item-level shifts rather than interest evolution; and (3) static feedback utilization, failing to incorporate dynamic user feedback during the recommendation. To address these challenges, we propose **ProRec-Video**, a proactive framework that guides hierarchical interest transitions through three innovations. First, *User Receptivity Profiling* assesses individual openness for new interests, ensuring personalized transition pacing. Second, *Hierarchical Interest Transition Planning* decomposes complex interest shifts into intermediate steps to generate smooth interest transition paths and semantically coherent video sequences, addressing overemphasis on item exposure. Third, *Dynamic Feedback Adaptation* integrates agent-based simulation and Reflexion mechanisms to refine interest transition paths and video sequences based on real-time user feedback, enhancing adaptability and satisfaction. Extensive experiments on two datasets demonstrate that ProRec-Video achieves a significant improvement in proactive recommendation performance, with an interest transition success rate of 85% and a user satisfaction rate of 78.3%.

Introduction

Short video platforms have revolutionized information consumption, with users generating and engaging with vast amounts of content daily. Traditional short video recommender systems (RSs) prioritize reinforcing users' existing preferences to maximize short-term engagement, often leading to information cocoons (Li et al. 2022) or opinion polarization (Bellina et al. 2023). While **proactive recommendations**(Zhu et al. 2023; Bi et al. 2024) have emerged as a promising approach to diversify content exposure and facilitate the expansion of interest. Unlike traditional RSs, proactive recommendations aim to guide users from their current



(a) Individual differences in acceptance of new interests

(b) Item-level rather than interest-level transformation

(c) Lack incorporation of user feedback for refinement

Figure 1: The illustration of the three principal limitations of existing proactive recommendation methods.

preferences to new and diverse domains. Recent studies have explored transformer-based methods (Zhu et al. 2023) and iterative frameworks (Bi et al. 2024) to enhance proactive recommendations. However, existing methods face three key limitations, as shown in Figure 1. (a) *Homogeneous Receptivity Assumption:* Most approaches assume uniform user receptivity to new interests (Zhu et al. 2023), ignoring individual differences in openness for novelty (e.g., some exploratory users prefer gradual exploration while others are conservative and resist abrupt shifts). During the proactive recommendation process, the negligence of the individual differences might lead to a low success rate of recommendations and user aversion. (b) *Short-term Item Exposure Without Interest Anchoring:* Current methods focus on increasing interaction probabilities with target items outside users' historical preferences but fail to build meaningful interest transitions (Bi et al. 2024). Although this approach may boost short-term engagement with the target item, it fails to in-

---

[*] Corresponding Author

duce long-term transformations in users' interests. (c) *Static Feedback Utilization:* User feedback is often overlooked or assumed to be passively accepted (Zhu et al. 2023), despite its critical role in refining proactive recommendations. In reality, users exhibit diverse reactions to different content, and that feedback can be considered as useful signals to improve the performance of proactive recommendation.

To address these shortcomings, we propose ProRec-Video, a proactive recommendation framework for short videos that integrates three innovations. First, **User Receptivity Profiling** quantifies individual tolerance for new interests by analyzing semantic differences between historical interactions, which helps establish a personalized pace of interest transition. Second, **Hierarchical Interest Transition Planning** shifts focus from item-level to interest-level guidance by decomposing complex interest shifts into structured intermediate steps via a two-stage pipeline, generating smooth interest transition paths and semantically coherent video sequences. This step leverages the rich knowledge and planning capabilities (Yang et al. 2024; Song et al. 2023) of Large Language Models (LLMs) to guide users from their current interests to new target areas and ensure semantic continuity. Third, **Dynamic Feedback Adaptation** integrates agent-based user simulation (Zhang et al. 2024) and Reflexion mechanisms (Shinn et al. 2024) to dynamically refine interest transition paths and video sequences based on real-time user feedback. This module enhances adaptability and satisfaction to ensure contextually relevant proactive recommendations. Extensive experimentation on two datasets demonstrates that ProRec-Video significantly improves the success rate of interest transitions and enhances user satisfaction, achieving an interest transition success rate of 85% and a user satisfaction rate of 78.3%. Our results prove the potential of ProRec-Video to advance proactive recommendations by enabling effective interest shifts and improving user experience.

The main contributions are summarized as follows:

- We propose ProRec-Video, a framework addressing three key limitations of existing proactive recommendations through user receptivity profiling, hierarchical interest transition planning, and dynamic feedback adaptation.
- We introduce a user receptivity profiling module to quantify individual novelty tolerance, facilitating personalized pacing of interest transitions.
- We design a hierarchical planning mechanism that shifts the focus from item to interest guidance, enabling smooth interest transition paths and coherent video sequences.
- We develop a feedback adaptation module that dynamically refines recommendations based on real-time user feedback, enhancing adaptability and satisfaction.
- We conduct extensive experiments on two datasets, demonstrating the effectiveness of ProRec-Video in improving interest transition success and user satisfaction.

Related Work

Short Video Recommendation

Short video recommendation poses unique challenges due to the passive nature of user interactions (Liu et al. 2024). Traditional sequential models (Wang et al. 2019), which primarily focus on explicit feedback such as clicks or purchases, fall short in capturing nuanced user preferences in this context. Pan et al. (Pan et al. 2023a) address this gap by incorporating passive-negative feedback alongside positive interactions, enhancing recommendation accuracy. Further advancements include optimizing implicit feedback signals in industrial settings. Specifically, Pan et al. (Pan et al. 2023b) propose a feedback-aware encoding module and multi-objective prediction mechanisms for platforms like Kuaishou (Gong et al. 2022), aiming to balance goals such as minimizing skipping rates and maximizing watch time. Additionally, Liu et al. (Liu et al. 2019) introduce an incremental multi-window scanning framework tailored for large-scale platforms like TikTok (Liu et al. 2022), designed to capture user interaction patterns efficiently while reducing computational complexity.

Proactive Recommendation

Proactive recommendation aims to guide users toward exploring new interest areas, mitigating issues such as filter bubbles (Wang et al. 2022) and opinion polarization (Bellina et al. 2023). A representative method is the Influential Recommender System (IRS) (Zhu et al. 2023), which constructs sequential influence paths to gradually increase the likelihood of user interaction with a target item. It employs a Transformer-based architecture with a Personalized Impressionability Mask (PIM) to model users' receptiveness to external influence. However, IRS lacks adaptability to real-time feedback and explicit guidance signals, limiting its effectiveness in dynamic environments. To address these limitations, Bi et al. (Bi et al. 2024) proposed Iterative Preference Guidance (IPG), a post-processing framework that iteratively refines user representations and re-ranks items based on interaction probability and guiding value. IPG explicitly incorporates guiding objectives without modifying the base RSs, offering greater flexibility across tasks. Nonetheless, both IRS and IPG focus on item-level guidance rather than interest-level transitions, which remains a key challenge in proactive recommendation.

Task Formulation

The goal of proactive recommendation for short videos is to guide users toward new interest domains while avoiding item-level myopia. Formally, let $u \in \mathcal{U}$ and $v \in \mathcal{V}$ denote a user and a video, respectively. Given the historical interaction $\mathcal{H}_u\{(v_i, t_i)\}$ of user u, where v_i represents a video watched by the user u, and t_i is the corresponding timestamp. The task is to construct a smooth interest transition path $\mathcal{P}_u = \{I_u \rightarrow I_{mid} \rightarrow I_{mid+1} \rightarrow \cdots \rightarrow I_t\}$ and a corresponding video sequence $\mathcal{S}_{\mathcal{P}_u} = \{v_1, v_2, \ldots, v_{|\mathcal{S}_{\mathcal{P}_u}|}\}$. The path $\mathcal{P}_u$ transitions users from their current interest I_u to a target interest I_t through intermediate interests aligned with semantic coherence. Moreover, the video sequence $\mathcal{S}_{\mathcal{P}_u}$ provides the specific videos to shift the user's interest from the current intermediate interest (e.g., I_{mid}) to the next interest (I_{mid+1}) through gradual semantic shifts and consolidates the newly introduced interest.

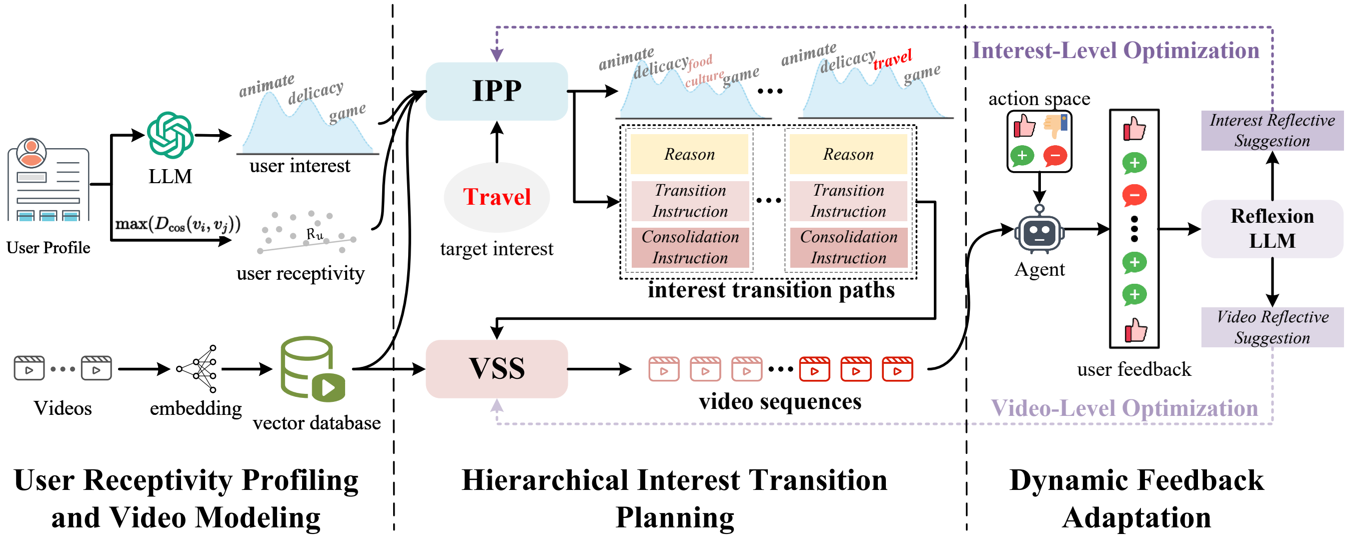


Figure 2: The overview of ProRec-Video. ProRec-Video includes three components: (1)User Receptivity and Video Modeling, (2) Hierarchical Interest Transition Planning, and (3) Dynamic Feedback Adaptation.

Methodology

The purpose of ProRec-Video is to transfer current interests of users to the target area as smoothly as possible. We illustrate ProRec-Video in Figure 2. In the User Receptivity Profiling and Video Modeling module, ProRec-Video constructs users' current interests and receptivity to new interests; meanwhile, it models the video embedding. The Hierarchical Interest Transition Planning module then analyzes this information and the target interest, generating a smooth interest transition path and coherent video sequences via a two-stage LLM pipeline. To improve the effectiveness of proactive recommendations and the experience of users, the path and sequences are updated iteratively based on the feedback in the Dynamic Feedback Adaptation module.

User Receptivity Profiling and Video Modeling

For user profiling, we focus on two key aspects: the user's current interest I_u and the user's receptivity R_u, which reflects their openness to new interests. Given that recent interactions are strong indicators of a user's current interest, we assign a time-based weight $\omega_i = exp^{-\lambda(t_{last}-t_i)}$ to each interaction v_i, where t_i denotes the timestamp of the interaction, t_{last} represents the most recent interaction timestamp, and λ is a time decay parameter. This temporal weighting produces a recent historical interaction sequence $\mathcal{H}_{recent}$, and we employ LLMs to infer the current interest I_u.

To model a user's tolerance for new interests, we introduce a user receptivity R_u. It is derived by segmenting the user's recent interaction sequence $\mathcal{H}_{recent}$ into sliding windows W of fixed length l and computing the semantic distance between the most divergent videos in each window:

$$D_w = \max_{i,j \in W}(1 - \cos(\mathbf{e}_{v_i}, \mathbf{e}_{v_j})), \quad W \subseteq \mathcal{H}_{recent}. \tag{1}$$

The user receptivity R_u is defined as the average of the maximum semantic distances across all windows:

$$R_u = \frac{1}{|W|} \sum_{W \subseteq \mathcal{H}_{recent}} D_w. \tag{2}$$

A higher R_u indicates greater willingness to new interests, while a lower value reflects a preference for familiar content.

This module forms the foundation of ProRec-Video by capturing video semantics and constructing user profiles that explicitly model current interests and individual receptivity levels to new interests. Each short video $v \in \mathcal{V}$ is represented as a multi-modal entity, consisting of visual, textual, and audio features. These features are encoded via modality-specific deep networks and fused through a cross-modal attention mechanism. The unified representation is projected into a semantic embedding space via $\mathbf{e}_v = EM(v)$, where $\mathbf{e}_v = EM(v)$ is a trainable neural encoder. The resulting embedding preserves high-level semantics (e.g., topic, sentiment) by aligning correlated modalities. This approach enables efficient similarity computation for retrieval and recommendation tasks. To facilitate efficient retrieval and mitigate recommendation hallucination (Perkovic, Drobnjak, and Boticki 2024), all semantic embeddings are stored in a vector database Σ.

Hierarchical Interest Transition Planning

This module addresses the challenge of item-level myopia by shifting focus from individual video recommendations to interest-level guidance. It generates smooth interest transition paths $\mathcal{P}_u$ and coherent video sequences $\mathcal{S}_{\mathcal{P}_u}$ through a two-stage hierarchical LLM pipeline, ensuring that recommendations align with the user's current interest I_u, target interest I_t, and user receptivity R_u. The first step focuses on the generation of a high-level user interest transition path using the **Interest Path Planner (IPP)**. After that, the second step involves the generation of specific short video sequences using the **Video Sequence Selector (VSS)**.

Interest Transition Path Generation First, IPP analyzes the user's current interest I_u and the target interest I_t, incorporating the user receptivity constraint R_u to ensure that consecutive interests along the transition path remain neither too broad nor too restrictive. Meanwhile, we apply Retrieval-Augmented Generation (RAG) (Lewis et al. 2020), leveraging contextual information from the vector database Σ to ensure that each intermediate interest is grounded in existing video content. This process can be formulated as:

$$\begin{aligned}\mathcal{P}_u &= \text{IPP}(I_u, I_t, R_u, \Sigma)\\ &= \{I_u \rightarrow I_{mid} \rightarrow I_{mid+1} \rightarrow \cdots \rightarrow I_t\}.\end{aligned} \tag{3}$$

Each pair of adjacent interests (e.g. $(I_{mid},\, I_{mid+1})$) within $\mathcal{P}_u$ must adhere to the user receptivity constraint in the semantic space: $D(I_{mid}, I_{mid+1}) < R_u$. This constraint ensures smooth transitions while maintaining user engagement and preventing abrupt shifts that may disrupt the user experience. To enhance the explainability of the interest transition and support the generation of video sequences, IPP provides detailed explanations and instructions for each transition $(I_{mid} \rightarrow I_{mid+1})$:

- **Reason** R_{mid+1}: The reason why the transition between the interest I_{mid} and the interest I_{mid+1} is reasonable.
- **Transition Instruction** TI_{mid+1}: An instruction designed to shift the user's interest from I_{mid} to I_{mid+1}.
- **Consolidation Instruction** CI_{mid+1}: An instruction to solidify the engagement of the user with I_{mid+1}.

Thus, the user interest transition path is structured as:

$$\begin{aligned}\mathcal{P}_u = \{I_u &\rightarrow (I_{mid}, R_{mid}, TI_{mid}, CI_{mid})\\ &\rightarrow (I_{mid+1}, R_{mid+1}, TI_{mid+1}, CI_{mid+1})\\ &\rightarrow \cdots \rightarrow (I_t, R_t, TI_t, CI_t)\}.\end{aligned} \tag{4}$$

For instance, consider an interest transition path involving the intermediate interests "nutrition tips" (I_{mid}) and "healthy meal recipes" (I_{mid+1}). In this case:

- R_{mid+1}: "The transition from 'nutrition tips' to 'healthy eating recipes' is reasonable, as both relate to food, with an emphasis shift towards health-conscious choices."
- TI_{mid+1}: "Recommend videos on preparing low-calorie or high-protein meals to guide the user's focus subtly."
- CI_{mid+1}: "Recommend videos on cooking healthy recipes to reinforce user interest."

Video Sequence Generation The next step translates the interest transition path $\mathcal{P}_u$ into specific video sequences $\mathcal{S}_{\mathcal{P}_u}$ by combining the transition instruction and the consolidation instruction. For each pair of adjacent interests $(I_{mid},\, I_{mid+1})$ in $\mathcal{P}_u$, the reason R_{mid+1}, transfer instruction TI_{mid+1}, and consolidation instruction CI_{mid+1} are input into VSS. These elements, along with contextual data retrieved from the video vector database Σ, allow VSS to interpret recommendation instructions more effectively and identify the most relevant videos. The context-aware constraints ensure that the generated short video sequences are semantically aligned with the intended interest transitions and meet user expectations. Corresponding video sequences for the transfer instruction TI_{mid+1} and consolidation instruction CI_{mid+1} are generated as follows:

$$\mathcal{S}_{TI_{mid+1}} = \text{VSS}(R_{mid+1},\, TI_{mid+1},\, \Sigma), \tag{5}$$

$$\mathcal{S}_{CI_{mid+1}} = \text{VSS}(R_{mid+1},\, CI_{mid+1},\, \Sigma). \tag{6}$$

Thus, the video sequence associated with the adjacent interest $(I_{mid},\, I_{mid+1})$ is formulated as:

$$\mathcal{S}_{mid+1} = \mathcal{S}_{TI_{mid+1}} \cup \mathcal{S}_{CI_{mid+1}}. \tag{7}$$

Finally, the complete video sequence $\mathcal{S}_{\mathcal{P}_u}$ for the interest transition path $\mathcal{P}_u$ is defined as:

$$\mathcal{S}_{\mathcal{P}_u} = \{\mathcal{S}_u,\, \mathcal{S}_{mid},\, \mathcal{S}_{mid+1},\, \ldots,\, \mathcal{S}_t\}. \tag{8}$$

Dynamic Feedback Adaptation

To mitigate the lack of user feedback engagement, we integrate an agent-based simulation and the Reflexion mechanism to optimize the interest transition path $\mathcal{P}_u$ and corresponding video sequences $\mathcal{S}_{\mathcal{P}_u}$. This module operates as a feedback loop with the hierarchical planning module, ensuring that $\mathcal{P}_u$ and $\mathcal{S}_{\mathcal{P}_u}$ adapt to user preferences and engagement patterns over time. The agent-based simulation makes use of the powerful capability of LLMs to approximate human reactions, simulating user positive and negative feedback during video consumption as optimization signals. Positive feedback includes *likes* and *positive comments*, while negative feedback comprises *dislikes* and *negative comments*. The agent provides feedback f_i for each video v_i in the video sequence $\mathcal{S}_{\mathcal{P}_u}$ after watching, where $f_i \in \{\textit{like},\ \textit{positive comment},\ \textit{dislike},\ \textit{negative comment}\}$.

To incorporate user feedback into the optimization of interest transition paths $\mathcal{P}_u$ and video sequences $\mathcal{S}_{\mathcal{P}_u}$, we employ the Reflexion mechanism, which operates at two levels: **interest-level optimization** and **video-level optimization**. In the interest-level optimization, the Reflexion LLM monitors the overall feedback $F = \{f_1, f_2, \ldots, f_{|\mathcal{S}_{\mathcal{P}_u}|}\}$ on the interest transition path $\mathcal{P}_u$. If a particular transition $(I_i \rightarrow I_{i+1})$ receives a high proportion of negative feedback (e.g., indicating a mismatch between adjacent interests), the Reflexion LLM will generate an interest-level reflective suggestion. After receiving this suggestion, IPP would regenerate a new interest transition path:

$$\mathcal{P}_u^{new} = \text{IPP}(I_u, I_t, R_u, \Sigma, \text{Ref}(\mathcal{P}_u|F)). \tag{9}$$

This process iterates until the feedback indicates that the transition path at the interest level is satisfactory or the maximum number of reflections is reached.

At the video level, if negative feedback is concentrated on a specific video v_i, the Reflexion LLM will generate a video-level reflective suggestion for VSS. VSS adjusts the video sequence based on this suggestion and ensures that each video maximizes user engagement while preserving the semantic alignment with the intermediate interest in $\mathcal{P}_u^{new}$. The reflection process continues iteratively until the Reflexion LLM ceases generating video-level refinement suggestions or the maximum refinement limit is reached. The combined interest-level and video-level reflection creates a closed-loop feedback system that dynamically adapts interest transition paths $\mathcal{P}_u$ and video sequences $\mathcal{S}_{\mathcal{P}_u}$.

	Methods	KuaiRec-Small					MicroLens-50K				
		SR	IoS	CO	GE	UES	SR	IoS	CO	GE	UES
Item-Level	Pf2Inf-Dijkstra	**0.5609**	-0.0822	0.1627	0.2478	0.4463	0.7000	-972.6250	0.2611	1.3659	0.7094
	Pf2Inf-MST	0.3314	-0.0356	0.1965	0.7089	0.4728	0.5385	-452.3846	0.2627	3.2806	0.6351
	Rec2Inf-GRU4Rec	0.1206	-3.6368	0.2607	6.2333	0.4503	0.5634	1293.4085	0.2462	4.7887	0.6818
	Rec2Inf-SASRec	0.1075	1.5648	0.2577	6.2487	0.4099	0.6714	1190.4429	0.2263	3.8577	0.7330
	Rec2Inf-TransRec	0.1619	-9.9391	0.2535	6.5742	0.4600	0.6615	**2099.6000**	0.1946	3.4278	0.7376
	IRS	0.2073	0.7091	0.2503	3.2202	0.3986	0.6667	761.1333	0.2813	3.2203	**0.7683**
	IPG	0.0513	-0.9060	0.2515	**13.2236**	0.4752	0.4074	16.4074	0.2859	9.0632	0.7208
	ProRec-Video	0.4561	**3.7160**	**0.6348**	9.9295	**0.6922**	**0.7222**	2093.5833	**0.5638**	**9.9740**	0.7513
Interest-Level	Pf2Inf-Dijkstra	0.5780	-1.1191	0.1570	0.2360	0.4542	0.6343	-862.9306	0.2616	1.3719	0.7493
	Pf2Inf-MST	0.3532	-1.5639	0.1957	0.7111	0.4904	0.5630	-599.0796	0.2699	3.4458	0.6358
	Rec2Inf-GRU4Rec	0.1651	4.5692	0.2597	6.0442	0.4583	0.4538	1198.1003	0.2539	4.8805	0.5828
	Rec2Inf-SASRec	0.1136	6.4663	0.2577	6.7678	0.4202	0.5487	1606.9502	0.2031	4.3688	0.6957
	Rec2Inf-TransRec	0.1158	-1.2776	0.2586	6.5108	0.4739	0.6988	1191.6565	0.2206	2.7979	0.6956
	IRS	0.2604	0.0396	0.2425	2.9758	0.4055	0.5720	630.8544	0.2966	3.2788	0.6740
	IPG	0.0197	3.7802	0.2504	**13.2222**	0.4727	0.4848	784.9659	0.2903	**8.7358**	0.7108
	ProRec-Video	**0.6860**	**29.7500**	**0.6302**	9.7329	**0.6698**	**0.8500**	**3351.4000**	**0.5585**	8.5819	**0.7830**

Table 1: The proactive recommendation performance between KuaiRec-Small and MicroLens-50K. Bold values indicate the best scores, while underlined values indicate the second best scores.

Experiment

Experimental Setup

Dataset We evaluate our method on two real-world short video datasets: **KuaiRec-Small** (Gao et al. 2022) and **MicroLens-50K** (Ni et al. 2023). KuaiRec-Small is derived from Kuaishou. It consists of 1,411 users and 3,327 videos, with a density of 99.6%. MicroLens-50K is a content-driven dataset comprising 34 million users and 1 million videos.

Simulation Evaluator Offline evaluation in proactive recommendation is challenging due to the difficulty of obtaining real-time ground truth on users' internal preferences and interest shifts. To overcome this, we adapt the Agent4Rec framework (Zhang et al. 2024) to simulate users in the short video domain. This modified framework generates feedback f_i for each video v_i based on users' current interests and evaluates whether users' interests have shifted toward the target interest using their feedback.

Hyperparameter We set the time decay parameter λ to 10^{-4} for KuaiRec-Small and 10^{-8} for MicroLens-50K. The maximum number of reflections is fixed at 3. For Rec2Inf, the top-k values are set to 20 and 100 on the two datasets, respecively. For the recommendation models (BERT4Rec (Sun et al. 2019), GRU4Rec (Tan, Xu, and Liu 2016), SASRec (Kang and McAuley 2018), TransRec (He, Kang, and McAuley 2017)), we use the default hyperparameters from Recbole (Xu et al. 2023).

Baseline We compare our approach with four baselines: Pf2Inf, Rec2Inf as outlined in (Zhu et al. 2023), and two proactive recommendation approaches, IRS and IPG.

- **Pf2Inf:** treats the interaction history as an undirected graph and applies path-finding algorithms to identify the shortest path from the user's current interest to the target item. We implement it with two algorithms: Dijkstra (Frana and Misa 2010) and MST (Bertsimas 1990).
- **Rec2Inf:** leverages RSs to generate influence paths through greedy search. It firstly selects k candidates from RSs and then appends items closest to the target until the target inclusion or reaching maximum length. We implement it with GRU4Rec, SASRec, and TransRec.
- **IRS:** employs a transformer-based influential network to model item sequence dependencies, along with PIM to capture user-specific susceptibility to external influences. Then IRS generates the influence path for each user.
- **IPG:** iteratively updates user representations and the interaction probability to generate candidates. In each iteration, IPG incorporates a guiding score to steer users toward the target item based on real-time feedback.

All four baselines originally focus on item-level targets. For fair comparison, we set the target of ProRec-Video to a specific video. For interest-level evaluation, we compute the average results of items within the same video category.

Metric To evaluate the performance of proactive recommendation methods, we employ the following five metrics:

- **Success Rate** (*SR*) quantifies the proportion that successfully lead the user u to watch the target video v_t:

$$SR = \frac{1}{|U|}\sum_{u=1}^{U} 1 \ \ (\text{if} \ \ u \text{ watch } v_t). \tag{10}$$

- **Increase of Score** (*IoS*) measures the improvement in the score of the target video v_t in the candidate corpus after user u watches the video sequence $\mathcal{S}_{\mathcal{P}_u}$:

$$IoS = \frac{1}{|U|}\sum_{u=1}^{|U|}(\text{Score}(v_t|\mathcal{H}_u \cup \mathcal{S}_{\mathcal{P}_u}) - \text{Score}(v_t|\mathcal{H}_u)), \tag{11}$$

where $\text{Score}(v|\mathcal{H}_u)$ denotes the predicted score of the video v, based on the historical interaction $\mathcal{H}_u$ of user u using Bert4Rec.

	Methods	KuaiRec-Small					MicroLens-50K				
		SR	IoS	CO	GE	UES	SR	IoS	CO	GE	UES
Resistant Group	Pf2Inf-Dijkstra	0.2979	5.1903	0.1650	0.2543	0.3759	0.5000	-1444.7500	0.2452	1.2091	0.7417
	Pf2Inf-MST	0.1111	6.0563	0.2026	0.7005	0.4636	0.1667	-1787.8333	0.2620	2.6408	0.4389
	Rec2Inf-GRU4Rec	0.0000	-2.1111	0.2571	6.2956	0.4522	0.2500	-381.8750	0.2565	3.3289	0.2875
	Rec2Inf-SASRec	0.0213	5.8671	0.2590	5.6219	0.3587	0.5000	1720.8333	0.1531	3.4116	0.5119
	Rec2Inf-TransRec	0.0286	-2.6136	0.2611	7.3582	0.3781	0.5000	-297.0000	0.1373	1.6824	0.5800
	IRS	0.0800	3.9015	0.2530	2.8152	0.3076	0.0000	308.5000	0.2988	2.9008	0.4375
	IPG	0.0000	2.0000	0.2468	**13.1671**	0.4306	0.1429	2241.8571	0.3960	**16.0031**	0.7579
	ProRec-Video	**0.3333**	**7.2205**	**0.6238**	9.0199	**0.7712**	**0.5833**	**4424.5000**	**0.5284**	8.8938	**0.7982**
Middle Group	Pf2Inf-Dijkstra	0.5748	-0.2827	0.1624	0.2480	0.4536	0.6970	-1075.0303	0.2600	1.3436	0.7036
	Pf2Inf-MST	0.3524	-0.4281	0.1954	0.7081	0.4735	0.5577	-172.5577	0.2594	3.3814	0.6482
	Rec2Inf-GRU4Rec	0.1442	0.2430	0.2611	6.2817	0.4506	0.6364	1798.9636	0.2450	4.8920	0.7518
	Rec2Inf-SASRec	0.1154	1.6941	0.2576	6.3886	0.4068	0.6964	1487.7857	0.2301	3.6231	0.7525
	Rec2Inf-TransRec	0.1680	**6.7204**	0.2528	6.5451	0.4679	0.6792	2608.6981	0.1956	3.5900	0.7592
	IRS	0.2153	0.2311	0.2487	3.2164	0.3999	0.7027	1155.1351	0.2775	3.4520	**0.7981**
	IPG	0.0612	-0.1531	0.2522	**13.2631**	0.4845	0.4615	-279.0769	0.2786	8.7234	0.7163
	ProRec-Video	**0.5855**	3.7016	**0.6313**	9.6973	**0.6629**	**0.8354**	**2983.1772**	**0.5592**	**9.0479**	0.7760
Willing Group	Pf2Inf-Dijkstra	0.7083	-2.8903	0.1631	0.2399	0.4566	**1.0000**	783.3333	0.2940	1.8202	0.7296
	Pf2Inf-MST	0.3922	-2.0910	0.1992	0.7247	0.4775	0.7143	-1386.4286	0.2876	3.0806	0.7061
	Rec2Inf-GRU4Rec	0.0732	-9.1596	0.2613	5.7974	0.4459	0.3750	-507.0000	0.2441	5.5380	0.5950
	Rec2Inf-SASRec	0.1463	-4.5715	0.2570	5.9020	0.4922	0.6250	-1288.7500	0.2548	5.8348	0.7617
	Rec2Inf-TransRec	0.2667	-3.2799	0.2506	5.9025	0.4896	0.5714	-43.1429	0.2281	3.4460	0.6864
	IRS	0.2826	-0.7026	0.2600	3.6908	0.4872	**1.0000**	-2430.7500	0.2990	1.3961	0.8244
	IPG	0.1200	-10.9000	0.2488	**12.8869**	0.4243	0.3750	-490.3750	0.2248	4.6474	0.7104
	ProRec-Video	**0.7143**	**2.6914**	**0.6517**	11.8408	**0.7539**	**1.0000**	**1063.0769**	**0.5851**	**10.1874**	**0.8515**

Table 2: Comparison of the guiding effect of users with different levels of acceptance of new interests. Bold values indicate the best scores, while underlined values indicate the second best scores.

- **Coherence** (*CO*) evaluates the semantic relevance between two consecutive videos (v_j, v_{j+1}) in $\mathcal{S}_{\mathcal{P}_u}$ by calculating the mean cosine similarity in the semantic space:

$$CO = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{1}{|\mathcal{S}_{\mathcal{P}_u}|} \sum_{v_j, v_{j+1} \in \mathcal{S}_{\mathcal{P}_u}} \cos\left(v_j, v_{j+1}\right). \quad (12)$$

- **Guiding Efficiency** (*GE*) quantifies the efficiency of the guided path by calculating the ratio of the semantic distance between v_j and v_{j+1} to the user receptivity:

$$GE = \frac{1}{|U|} \sum_{u=1}^{|U|} \frac{1}{|\mathcal{S}_{\mathcal{P}_u}|} \sum_{v_j, v_{j+1} \in \mathcal{S}_{\mathcal{P}_u}} \frac{1 - \cos(v_j, v_{j+1})}{R_u}. \quad (13)$$

- **User Experience Score** (*UES*) assesses user satisfaction throughout the guidance process:

$$UE = \frac{1}{|U|} \sum_{u}^{U} \frac{n_{pos}}{n_{pos} + n_{neg}}, \quad (14)$$

where n_{pos} and n_{neg} denote the number of positive and negative feedback instances, respectively.

Main Results

We evaluate ProRec-Video against baselines on KuaiRec-Small and MicroLens-50K for item-level and interest-level proactive recommendation. As shown in Table 1, ProRec-Video demonstrates super IoS performance across both tasks, with particularly strong results at the interest level.

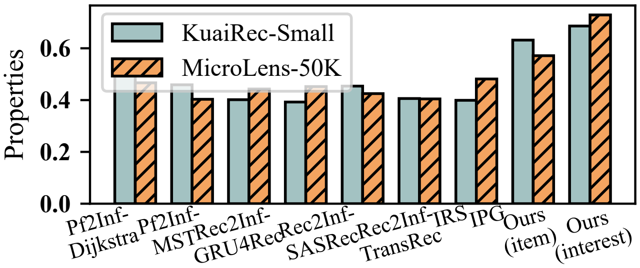


Figure 3: Comparison of item and interest-level methods.

Item-Level Results ProRec-Video achieves a success rate of 0.4561 on KuaiRec-Small, slightly lower than Pf2Inf-Dijkstra due to the dataset's dense connectivity. However, it shows better generalization on the sparser MicroLens-50K. While Pf2Inf achieves higher SR on KuaiRec-Small, it shows minimal improvement in IoS, whereas ProRec-Video consistently outperforms in this metric. ProRec-Video also achieves the highest coherence and guiding efficiency on MicroLens-50K, and maintains a user experience score of nearly 70%, outperforming most baselines.

Interest-Level Results ProRec-Video remarkably outperforms all baselines, achieving a SR of 0.6859 on KuaiRec-Small and 0.85 on MicroLens-50K. It also achieves the highest IoS, up to four times greater than Rec2Inf-SASRec on KuaiRec-Small. ProRec-Video achieves coherence scores of

feedback	KuaiRec-Small						MicroLens-50K					
	SR	IoS	CO	GE	UES	Ref_Number	SR	IoS	CO	GE	UES	Ref_Number
none	0.6634	8.5000	0.6325	**10.6824**	0.6654	—	0.6939	1883.3878	0.5523	**10.0643**	0.7365	—
positive	0.6000	-22.6667	0.6204	9.1201	0.6473	2.7400	0.7500	2003.6383	0.5476	8.4272	0.7535	1.9773
negative	0.6207	-37.0000	0.5993	8.4453	0.6467	2.7931	0.7832	1321.2955	0.5374	8.3137	0.7830	2.3404
both	**0.6860**	**29.7500**	**0.6602**	9.7329	**0.6998**	1.3140	**0.8500**	**3351.4000**	**0.5685**	8.5819	**0.7880**	1.8000

Table 3: The impact of different feedback mechanisms on proactive recommendation effectiveness.

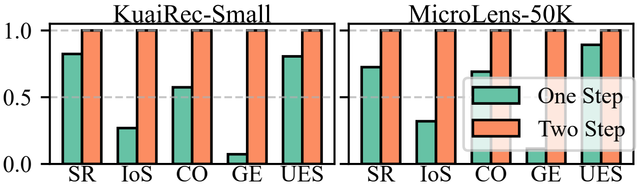


Figure 4: Impact of one-step vs. two-step LLM planning.

0.6302 and 0.5585 on the two datasets, respectively, indicating super IoS semantic smoothness. UES remains consistently above 70%, outperforming baselines by 10–20%.

Interest Shifting Analysis To verify whether users' interests shift toward the target I_t, we update user interest representations following Section and compare them with the target interest. As shown in Figure 3, ProRec-Video with interest-level guidance achieves significantly higher SR in interest transition than item-level methods, demonstrating its effectiveness in guiding long-term interest evolution.

Various User Groups Analysis

User receptivity to new interests is different. To investigate the impact of various user receptivity on proactive recommendation, we categorize users into three groups based on receptivity: **Resistant** (bottom 10%), **Middle** (80%), and **Willing** (top 10%). As shown in Table 2, ProRec-Video outperforms all baselines across all groups on both datasets.

Resistant Group Results Guiding users in the Resistant Group is particularly challenging due to their low openness to new content. Several baselines (e.g., Rec2Inf and IPG on KuaiRec-Small, IRS on MicroLens-50K) perform near zero in success rate. In contrast, ProRec-Video achieves the highest SR, IoS, CO, and UES, demonstrating its effectiveness in guiding even the most resistant users toward new interests.

Middle Group Results Middle users are moderately receptive. ProRec-Video still outperforms all baselines. It achieves a SR of 0.5855 on KuaiRec-Small and 0.8354 on MicroLens-50K. Its super IoS coherence indicates smooth and semantically relevant video sequences, which are crucial for effective interest transition and user engagement.

Willing Group Results In the Willing Group, all methods perform reasonably well. However, ProRec-Video still achieves the highest SR, along with super IoS IoS, CO, and UES. These results confirm its robustness and effectiveness even in the most favourable user conditions.

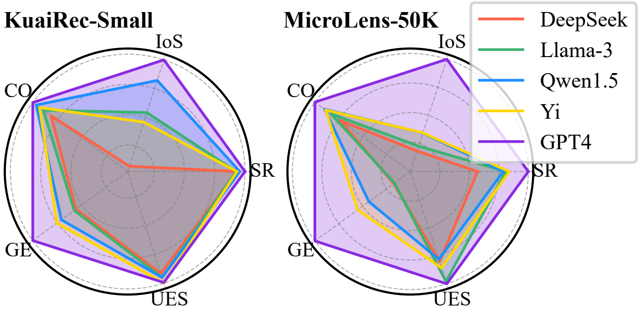


Figure 5: Comparison of different LLMs as IPP and VSS.

Ablation Studies

Single vs. Two-Step LLMs To evaluate the impact of two-stage LLM pipeline, we merge it into a single process and compare it with the two-step process. Results in Figure 4 indicate that a single-step process significantly reduces every metric. It suggests that separating the proactive recommendation into distinct stages enhances overall performance.

Impact of Different LLMs To assess the impact of different LLMs as IPP and VSS, we select DeepSeek-7B (DeepSeek-AI 2024), Llama-3-8B (Dubey et al. 2024), Qwen1.5-32B (Bai et al. 2023), and Yi-34B (AI et al. 2024). Experimental results in Figure 5 show that LLMs with larger parameters outperform smaller models. Notably, GPT-4 achieves the best performance.

Feedback Optimization To analyze the impact of feedback, we set three configurations: no feedback (**none**), positive feedback only (**positive**), and negative feedback only (**negative**). Results at Table 3 demonstrate that integrating **both** types of feedback yields the best performance across most metrics. Using only one type of feedback is less effective than none, and increases the cost of reflection.

Conclusion

In this paper, we propose ProRec-Video, a proactive recommendation framework that enables user interest evolution through personalised user receptivity profiling, hierarchical interest transition planning, and dynamic feedback adaptation. Experimental results demonstrate that ProRec-Video consistently outperforms state-of-the-art proactive recommendation methods across key metrics on two short video recommendation datasets.

References

AI, .; :; Young, A.; Chen, B.; Li, C.; Huang, C.; Zhang, G.; Zhang, G.; Li, H.; Zhu, J.; Chen, J.; Chang, J.; Yu, K.; Liu, P.; Liu, Q.; Yue, S.; Yang, S.; Yang, S.; Yu, T.; Xie, W.; Huang, W.; Hu, X.; Ren, X.; Niu, X.; Nie, P.; Xu, Y.; Liu, Y.; Wang, Y.; Cai, Y.; Gu, Z.; Liu, Z.; and Dai, Z. 2024. Yi: Open Foundation Models by 01.AI. arXiv:2403.04652.

Bai, J.; Bai, S.; Chu, Y.; Cui, Z.; Dang, K.; Deng, X.; Fan, Y.; Ge, W.; Han, Y.; Huang, F.; et al. 2023. Qwen technical report. *arXiv preprint arXiv:2309.16609*, abs/2309.16609.

Bellina, A.; Castellano, C.; Pineau, P.; Iannelli, G.; and De Marzo, G. 2023. Effect of collaborative-filtering-based recommendation algorithms on opinion polarization. *Phys. Rev. E*, 108: 054304.

Bertsimas, D. J. 1990. The probabilistic minimum spanning tree problem. *Networks*, 20(3): 245–275.

Bi, S.; Wang, W.; Pan, H.; Feng, F.; and He, X. 2024. Proactive Recommendation with Iterative Preference Guidance. In Chua, T.; Ngo, C.; Lee, R. K.; Kumar, R.; and Lauw, H. W., eds., *Companion Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, Singapore, May 13-17, 2024*, 871–874. Singapore: ACM.

DeepSeek-AI. 2024. DeepSeek LLM: Scaling Open-Source Language Models with Longtermism. *arXiv preprint arXiv:2401.02954*, abs/2401.02954.

Dubey, A.; Jauhri, A.; Pandey, A.; and et al. 2024. The Llama 3 Herd of Models. *CoRR*, abs/2407.21783.

Frana, P. L.; and Misa, T. J. 2010. An interview with edsger w. dijkstra. *Communications of the ACM*, 53(8): 41–47.

Gao, C.; Li, S.; Lei, W.; Chen, J.; Li, B.; Jiang, P.; He, X.; Mao, J.; and Chua, T.-S. 2022. KuaiRec: A Fully-Observed Dataset and Insights for Evaluating Recommender Systems. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, CIKM '22, 540–550. Atlanta, GA, USA: Association for Computing Machinery.

Gong, X.; Feng, Q.; Zhang, Y.; Qin, J.; Ding, W.; Li, B.; Jiang, P.; and Gai, K. 2022. Real-time Short Video Recommendation on Mobile Devices. In Hasan, M. A.; and Xiong, L., eds., *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, 3103–3112. Atlanta, GA, USA: ACM.

He, R.; Kang, W.; and McAuley, J. J. 2017. Translation-based Recommendation. In Cremonesi, P.; Ricci, F.; Berkovsky, S.; and Tuzhilin, A., eds., *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017*, 161–169. Como, Italy: ACM.

Kang, W.; and McAuley, J. J. 2018. Self-Attentive Sequential Recommendation. In *IEEE International Conference on Data Mining, ICDM 2018, Singapore, November 17-20, 2018*, 197–206. Singapore: IEEE Computer Society.

Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.-t.; Rocktäschel, T.; et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33: 9459–9474.

Li, N.; Gao, C.; Piao, J.; Huang, X.; Yue, A.; Zhou, L.; Liao, Q.; and Li, Y. 2022. An Exploratory Study of Information Cocoon on Short-form Video Platform. In Hasan, M. A.; and Xiong, L., eds., *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*, 4178–4182. Atlanta, GA, USA: ACM.

Liu, Q.; Hu, J.; Xiao, Y.; Zhao, X.; Gao, J.; Wang, W.; Li, Q.; and Tang, J. 2024. Multimodal Recommender Systems: A Survey. *ACM Comput. Surv.*, 57(2).

Liu, Y.; Lyu, C.; Liu, Z.; and Tao, D. 2019. Building Effective Short Video Recommendation. In *IEEE International Conference on Multimedia & Expo Workshops, ICME Workshops 2019, Shanghai, China, July 8-12, 2019*, 651–656. Shanghai, China: IEEE.

Liu, Z.; Zou, L.; Zou, X.; Wang, C.; Zhang, B.; Tang, D.; Zhu, B.; Zhu, Y.; Wu, P.; Wang, K.; and Cheng, Y. 2022. Monolith: Real Time Recommendation System with Collisionless Embedding Table. In Vinagre, J.; Al-Ghossein, M.; Jorge, A. M.; Bifet, A.; and Peska, L., eds., *Proceedings of the 5th Workshop on Online Recommender Systems and User Modeling co-located with the 16th ACM Conference on Recommender Systems, ORSUM@RecSys 2022, Seattle, WA, USA, September 23rd, 2022*, volume 3303 of *CEUR Workshop Proceedings*, 64–72. Seattle, WA, USA: CEUR-WS.org.

Ni, Y.; Cheng, Y.; Liu, X.; Fu, J.; Li, Y.; He, X.; Zhang, Y.; and Yuan, F. 2023. A Content-Driven Micro-Video Recommendation Dataset at Scale. *CoRR*, abs/2309.15379.

Pan, Y.; Gao, C.; Chang, J.; Niu, Y.; Song, Y.; Gai, K.; Jin, D.; and Li, Y. 2023a. Understanding and Modeling Passive-Negative Feedback for Short-video Sequential Recommendation. In Zhang, J.; Chen, L.; Berkovsky, S.; Zhang, M.; Noia, T. D.; Basilico, J.; Pizzato, L.; and Song, Y., eds., *Proceedings of the 17th ACM Conference on Recommender Systems, RecSys 2023, Singapore, Singapore, September 18-22, 2023*, 540–550. Singapore: ACM.

Pan, Y.; Li, N.; Gao, C.; Chang, J.; Niu, Y.; Song, Y.; Jin, D.; and Li, Y. 2023b. Learning and Optimization of Implicit Negative Feedback for Industrial Short-video Recommender System. In Frommholz, I.; Hopfgartner, F.; Lee, M.; Oakes, M.; Lalmas, M.; Zhang, M.; and Santos, R. L. T., eds., *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*, 4787–4793. Birmingham, United Kingdom: ACM.

Perkovic, G.; Drobnjak, A.; and Boticki, I. 2024. Hallucinations in LLMs: Understanding and Addressing Challenges. In Babic, S.; Car, Z.; Cicin-Sain, M.; Cisic, D.; Ergovic, P.; Grbac, T. G.; Gradisnik, V.; Gros, S.; Jokic, A.; Jovic, A.; Jurekovic, D.; Katulic, T.; Koricic, M.; Mornar, V.; Petrovic, J.; Skala, K.; Skvorc, D.; Sruk, V.; Svaco, M.; Tijan, E.; Vrcek, N.; and Vrdoljak, B., eds., *47th MIPRO ICT and Electronics Convention, MIPRO 2024, Opatija, Croatia, May 20-24, 2024*, 2084–2088. Opatija, Croatia: IEEE.

Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; and Yao, S. 2024. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36.

Song, C. H.; Sadler, B. M.; Wu, J.; Chao, W.; Washington, C.; and Su, Y. 2023. LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, 2986–2997. Paris, France: IEEE.

Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; and Jiang, P. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In Zhu, W.; Tao, D.; Cheng, X.; Cui, P.; Rundensteiner, E. A.; Carmel, D.; He, Q.; and Yu, J. X., eds., *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM 2019, Beijing, China, November 3-7, 2019*, 1441–1450. Beijing, China: ACM.

Tan, Y. K.; Xu, X.; and Liu, Y. 2016. Improved Recurrent Neural Networks for Session-based Recommendations. In Karatzoglou, A.; Hidasi, B.; Tikk, D.; Shalom, O. S.; Roitman, H.; Shapira, B.; and Rokach, L., eds., *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016*, 17–22. Boston, MA, USA: ACM.

Wang, S.; Hu, L.; Wang, Y.; Cao, L.; Sheng, Q. Z.; and Orgun, M. A. 2019. Sequential Recommender Systems: Challenges, Progress and Prospects. In Kraus, S., ed., *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 6332–6338. Macao, China: ijcai.org.

Wang, W.; Feng, F.; Nie, L.; and Chua, T. 2022. User-controllable Recommendation Against Filter Bubbles. In Amigó, E.; Castells, P.; Gonzalo, J.; Carterette, B.; Culpepper, J. S.; and Kazai, G., eds., *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, 1251–1261. Madrid, Spain: ACM.

Xu, L.; Tian, Z.; Zhang, G.; Zhang, J.; Wang, L.; Zheng, B.; Li, Y.; Tang, J.; Zhang, Z.; Hou, Y.; Pan, X.; Zhao, W. X.; Chen, X.; and Wen, J. 2023. Towards a More User-Friendly and Easy-to-Use Benchmark Library for Recommender Systems. In *SIGIR*, 2837–2847. Taipei, Taiwan: ACM.

Yang, J.; Jin, H.; Tang, R.; Han, X.; Feng, Q.; Jiang, H.; Zhong, S.; Yin, B.; and Hu, X. B. 2024. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. *ACM Trans. Knowl. Discov. Data*, 18(6): 160:1–160:32.

Zhang, A.; Chen, Y.; Sheng, L.; Wang, X.; and Chua, T. 2024. On Generative Agents in Recommendation. In Yang, G. H.; Wang, H.; Han, S.; Hauff, C.; Zuccon, G.; and Zhang, Y., eds., *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, 1807–1817. Washington DC, USA: ACM.

Zhu, H.; Ge, H.; Gu, X.; Zhao, P.; and Lee, D. L. 2023. Influential Recommender System. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023*, 1406–1419. Anaheim, CA, USA: IEEE.