Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

Latest updates: https://dl.acm.org/doi/10.1145/3715099

RESEARCH-ARTICLE

LLMCDSR: Enhancing Cross-Domain Sequential Recommendation with Large Language Models

HAORAN XIN, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China

YING SUN, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China

CHAO WANG, University of Science and Technology of China, Hefei, Anhui, China

HUI XIONG, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, Guangdong, China

Open Access Support provided by:

The Hong Kong University of Science and Technology (Guangzhou)

University of Science and Technology of China

PDF Download
3715099.pdf
07 April 2026
Total Citations: 16
Total Downloads: 3271

Published: 10 July 2025
Online AM: 28 January 2025
Accepted: 02 January 2025
Revised: 12 November 2024
Received: 02 February 2024

Citation in BibTeX format

LLMCDSR: Enhancing Cross-Domain Sequential Recommendation with Large Language Models

HAORAN XIN and YING SUN , Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China
CHAO WANG , School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei, China
HUI XIONG , Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China and Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China

Cross-Domain Sequential Recommendation (CDSR) aims to predict users' preferences based on historical sequential interactions across multiple domains. Existing works focus on the overlapped users who interact in multiple domains to capture the cross-domain correlations. These methods often underperform in practical scenarios featuring both overlapped and non-overlapped users due to the limited cross-domain interactions and knowledge transfer misalignment for non-overlapped users. To address this, we leverage Large Language Models (LLMs) to facilitate CDSR by fully exploiting single-domain interactions. However, LLMs exhibit inherent limitations in handling extensive item repositories and sequential collaborative signals. Moreover, the generation reliability is compromised by the hallucination problem, potentially causing noisy and unstable outputs. To this end, we propose a novel *LLMCDSR* framework, which employs LLMs to predict unobserved cross-domain interactions, termed pseudo items, within single-domain interactions. Specifically, we first prompt LLMs to execute the Candidate-Free Cross-Domain Interaction Generation task. Then, we devise a Collaborative-Textual Contrastive Pre-Training strategy, learning to infuse collaborative information into textual features. Afterwards, we present a novel Relevance-Aware Meta Recall Network (RMRN) to selectively identify and retrieve high-quality pseudo items from the dataset, where the parameters are optimized in a meta-learning manner. Finally, extensive experiments on two public datasets validate the effectiveness of LLMCDSR in enhancing CDSR. The code and data are available at https://github.com/xhran2010/LLMCDSR.
CCS Concepts: • **Information systems** → **Recommender systems**;

Additional Key Words and Phrases: Cross-domain recommendation, Sequential recommendation, Large language models

This work was supported in part by the National Natural Science Foundation of China (Grant No.92370204, 62306255), in part by the National Key R&D Program of China (Grant No.2023YFF0725001), in part by Guangzhou-HKUST(GZ) Joint Funding Program (Grant No.2023A03J0008), in part by the Education Bureau of Guangzhou Municipality, and in part by the Natural Science Foundation of Guangdong Province (No. 2024A1515011839).
Authors' Contact Information: Haoran Xin, Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; e-mail: hxin883@connect.hkust-gz.edu.cn; Ying Sun (corresponding author), Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; e-mail: yings@hkust-gz.edu.cn; Chao Wang, School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei, China; e-mail: wangchaoai@ustc.edu.cn; Hui Xiong (corresponding author), Thrust of Artificial Intelligence, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China and Department ofComputer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China; e-mail: xionghui@ust.hk.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 1558-2868/2025/7-ART120
https://doi.org/10.1145/3715099

ACM Reference format:
Haoran Xin, Ying Sun, Chao Wang, and Hui Xiong . 2025. LLMCDSR: Enhancing Cross-Domain Sequential Recommendation with Large Language Models. *ACM Trans. Inf. Syst.* 43, 5, Article 120 (July 2025), 33 pages. https://doi.org/10.1145/3715099

1 Introduction

Sequential Recommendation (SR) has become indispensable in personalizing user experiences on various platforms, leveraging the power of historical interaction patterns to predict future preferences. In fact, users often engage in multiple domains, and user behaviors across different domains often exhibit similar tendencies, implying that a user's interactions in one domain can provide valuable insights for another. Recognizing this interconnection, **Cross-Domain SR (CDSR)** aims to predict the interactions in multiple domains based on the user's cross-domain historical records.

Extensive research efforts [5, 23, 55, 56, 92] for CDSR are dedicated to learning the cross-domain sequential correlations relying on the overlapped users that are active in multiple domains. For example, Ma et al. [56] employed the gating mechanism to adaptively filter out domain-irrelevant information in the cross-domain sequences given an overlapped user. Cao et al. [5] devised a contrastive Infomax objective for the user's representations in different domains to capture the nuanced cross-domain correlations. Yet, in practical scenarios, the proportion of overlapped users is often restricted [41], potentially limiting the effectiveness of those methods. Moreover, as shown in Figure 1(a), there exists a significant population of non-overlapped users, who engage exclusively within a single domain and contribute a large volume of single-domain interactions. These interactions, rich in single-domain collaborative information, offer significant insights into behavioral patterns within their specific domain. Nevertheless, their potential for enhancing CDSR is hindered by the lack of cross-domain correlations, limiting their applicability in broader contexts.

Therefore, acknowledging the untapped potential in single-domain interactions of non-overlapped users, we aim to fully harness these interactions to enhance CDSR. Essentially, the underutilization of non-overlapped information primarily stems from a fundamental gap in the feature spaces of single- and cross-domain sequences. It is challenging to align these two data sources due to the scarcity of cross-domain interaction data, coupled with the absence of cross-domain indicators in single-domain data. This misalignment consequently impedes the effective transfer of preference knowledge across domains. With **Large Language Models (LLMs)**, such as ChatGPT [4] and Llama 2 [75], showing impressive world knowledge on a diversity of topics, they are acting as implicit knowledge bases and have proven to be powerful few-shot and even zero-shot learners [1, 54, 85]. As the cross-domain text is implicitly included in the web corpora (e.g., related books for a movie), LLMs have the potential to understand the cross-domain knowledge, thereby bridging the domain gap and aligning the single- and cross-domain data. However, leveraging LLMs for CDSR presents several challenges. Firstly, platforms typically have a broad and established set of items with distinct identifiers. Nevertheless, due to the contextual limitation of LLMs, they struggle to fully process this vast item repository and accurately match existing items. Secondly, LLMs are susceptible to the hallucination problem [99], which can result in noisy and unstable outputs that are difficult to evaluate and use effectively. Additionally, understanding users' cross-domain behaviors typically requires task-specific collaborative information, which is usually outside the common knowledge of LLMs. Nevertheless, it is hard to integrate these collaborative signals into LLM queries, especially when the signals possess sequentialcharacteristics.

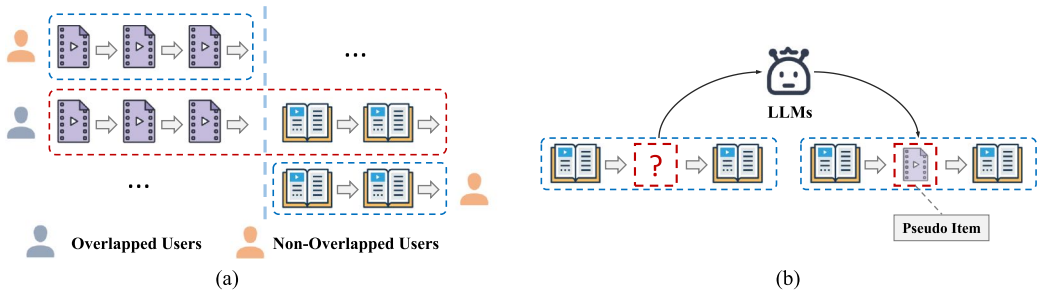


Fig. 1. (a) The illustration of the CDSR problem where both overlapped and non-overlapped users exist. (b) We leverage LLMs to predict pseudo-cross-domain items in single-domain sequences.

To this end, this paper proposes LLMCDSR, a novel framework to leverage LLMs for CDSR featuring both overlapped and non-overlapped users. In general, as users frequently engage in multiple domains, those exhibiting single-domain interactions also likely have unobserved activities in other domains. Acknowledging the integrated user profile and temporal dependencies between domain behaviors, as shown in Figure 1(b), we propose prompting LLMs to predict these latent cross-domain interactions, referred to as *pseudo items*, based on LLMs' extensive cross-domain knowledge. These pseudo items are effectively integrated into enhanced sequences by leveraging collaborative information and evaluating their relevance to the task. These steps together can improve domain alignment, enabling the inclusion of non-overlapped users to enrich CDSR. Specifically, to alleviate the burden on LLMs of processing the vast candidate space and complex collaborative information, we first design a prompt template to guide LLMs to accomplish the *Candidate-Free Cross-Domain Interaction Generation*. Then, we propose a *Collaborative-Textual Contrastive Pre-Training* strategy, learning to inject the task-specific collaborative features into the textual features of LLMs. Afterwards, we present a **Relevance-Aware Meta Recall Network (RMRN)** to selectively map the high-quality generated samples to the items of the dataset. Technically, it incorporates a novel relevance weighting layer, which adaptively assesses the relevance of LLM-generated samples w.r.t. the downstream recommendation task. This is followed by a recall operation, which retrieves the pseudo items in the dataset based on the collaborative similarity. Finally, motivated by the meta-learning schema, we introduce an advanced bi-level optimization to jointly train our RMRN as well as the downstream recommender, improving the relevance of the pseudo items conforming to the real cross-domain interactions.

The main contributions of this work are highlighted as follows:

- We study the CDSR problem where both overlapped and non-overlapped users present simultaneously, which is a challenging scenario due to the limited cross-domain knowledge alignment and transfer.
- To fully exploit the non-overlapped information for enhancing CDSR, we propose a novel framework, coined LLMCDSR, which leverages LLMs to predict pseudo-cross-domain items in single-domain sequences and incorporates an RMRNto selectively retrieve relevant pseudo items, thereby strengthening the knowledge alignment and transfer across domains.
- We design a contrastive pre-training strategy, learning to enrich the textual features with collaborative information. An advanced bi-level optimization based on meta-learning is further presented to improve the relevance of the pseudo items, aligning them more closely with actual cross-domain interactions.

—We conduct extensive experiments on two public real-world datasets to validate the effectiveness of our proposed LLMCDSR framework against the CDSR problem.

2 Related Works

In this section, we present a concise overview of related literature in three areas: (1) SR, (2) cross-domain recommendation, and (3) LLMs for recommendation.

2.1 SR

Recommender systems [44, 72, 79–81, 94, 97, 101] play an important role in people's daily life by learning and predicting user preferences through their interactions with items. Among users' behaviors, sequential interaction is pivotal in revealing the ever-changing preferences [10, 73]. Those widely employed traditional **Collaborative Filtering (CF)** methods [26, 27, 48, 82, 91] often stumble in this regard due to their inherent inability to capture sequential signals within user data. These limitations have aroused the study of SR approaches. Early endeavors in addressing this gap leveraged the Markov Decision Process [67] and **Markov Chains (MC)** [25, 27, 65] to model the sequential pattern of item sequences. However, these models focused on immediate interactions, thus neglecting the broader historical context that often shapes user comprehensive preferences.

Afterwards, a number of deep learning-based models [16, 28, 60, 83] that incorporate **Recurrent Neural Networks (RNNs)** [11, 29] have been proposed to tackle the SR problem, benefiting from the advancement of RNNs in capturing the length-variable sequential interactions. Hidasi et al. [28] is the early work that employed **Gated Recurrent Unit (GRU)** [11], a variant of RNNs, to model the interaction sessions by introducing the session-parallel training strategy. Donkers et al. [16] redesigned the gate mechanism in the GRU cell to incorporate fine-grained user and item information. Besides, the potential of neural networks for handling other data types in modeling users' interaction sequences has also been investigated. For example, Tang and Wang [74] proposed to transform users' sequences into an "image,", with the Convolutional Neural Networks employed to identify sequential patterns as local features within this so-called image. Huang et al. [32] utilized convolutional-recurrent encoders to extract patterns of user behavior in terms of both category and time across time scales. Several works [53, 88, 98] also employed **Graph Neural Networks (GNNs)** [38, 78] to capture the complex transition patterns lying in the sequences beyond 1-D temporal patterns by treating the item transitions as graph-structured data.

Recently, as the self-attention mechanism [77] has demonstrated impressive effectiveness across several natural language processing [13] and computer vision tasks [52], it has also been applied in sequential recommender systems [6, 17, 18, 24, 36] for capturing the higher-order dependencies in the sequence, and achieved better performance. For example, SASRec [36] was the early attempt to apply the self-attention mechanism to SR problems, which is able to adaptively assign weights to previous items, and the efficiency was improved owing to the parallel acceleration of such architecture. Du et al. [17] probed the self-attention mechanism from the perspective of the frequency domain where the periodic characteristics are addressed and the time and frequency features are fused in a union model.

In this article, we engage in the cross-domain recommendation study based on the SR foundations. We study how to improve the SR performance across multiple domains with the utilization of both overlapped and non-overlapped users.

2.2 Cross-Domain Recommendation

The cross-domain recommendation aims to overcome the limitations of single-domain recommender systems, such as data sparsity and the cold-start problem. A large number of works focus on the general cross-domain recommendation task based on CF. As Zang et al. [96] suggest, these works

can be categorized into three main types, i.e., factorization-based, feature-based, and transfer-based methods. Factorization-based methods leverage factorization techniques to extract and utilize latent features from complex user-item interactions. Several works [34, 62, 69] extended traditional matrix factorization for cross-domain recommendation by decomposing the domain-specific rating matrix into user and item latent features, with the cross-domain knowledge guiding such the process. Moving forward, more fine-grained factors [31, 50] such as time and aspect were also incorporated into the factorization procedure. Feature-based methods leverage user or item embeddings to share knowledge across domains. For example, Zhao et al. [105] proposed a multi-sparse-domain collaborative recommendation method to enhance the user's domain-specific preference by being shared with the user's complementary aspect preferences in other domains. Transfer-based methods aim at learning the knowledge transfer across domains. A series of works [39, 57, 106, 107] utilized the embedding and mapping paradigm to map the features from the source domain to the target domain. Among these works, EMCDR [57] is one of the representative works that first learned the latent features in respective domains and then employed a multi-layer perceptron to learn the non-linear projection across domains. Several works [30, 47, 49, 104] were also proposed to learn the transferability across domains simultaneously in an end-to-end manner. For example, Hu et al. [30] first proposed a collaborative cross-network that introduced cross-connections from one base network to another. Liu et al. [49] further incorporated GNNs to fulfill the dual knowledge transfer based on a user-item bipartite graph in each domain. However, although these methods have remarkable performance on general cross-domain tasks, they cannot be directly applied to the SR due to the shortage of sequential signal awareness.

Recently, some pioneer works have also explored the CDSR problem. Ma et al. [56] and Sun et al. [70] focused on the shared-account scenario and employed the gating mechanism to filter out domain-irrelevant information of overlapped users. Some works [23, 55] used the power of knowledge graphs to enhance knowledge transfer across domains. Besides, several self-supervised approaches [5, 84, 95] have been proposed to capture the nuanced cross-domain correlations. Nevertheless, these methods largely depend on the overlapped users for knowledge transfer, which is not always practical in real-world scenarios. Lin et al. [43] introduced the group-level representation to make use of the information of non-overlapped users without the constraint of overlapping. However, the sequences in respective domains are separately learned and inferred, potentially limiting the approach's scalability and applicability.

In this article, we focus on the more general scenario in CDSR where both overlapped and non-overlapped users present and alleviate the dependency on overlapped users by estimating the latent pseudo cross-domain sequences based on the original sequences from non-overlapped users, thereby the abundant information lying in the non-overlapped users can be fully utilized to facilitate the knowledge alignment and transfer across domains.

2.3 LLMs for Recommendation

Recently, the advent of LLMs in natural language processing [7, 103] has sparked increased interest in leveraging these models to improve recommender systems. Owing to the ability of LLMs to process and comprehend human language, many content-based recommender systems Zhang et al. [100] have the potential to benefit from the power of LLMs. In the early attempts, the causal language models (e.g., BERT [13]) have been utilized as the textual encoders to encode the textual side information of users and items because of their remarkable discriminative ability. The encoded textual representations are usually integrated into the embeddings of the downstream recommendation models to promote recommendation performance. For example, Qiu et al. [59] and Wu et al. [87] fine-tuned BERT with the textual information from content-rich domains, and learned the user representation to improve the recommendation in domains with sparse information.

The prompt learning technique has also been explored with the causal language models on the recommendation task [19, 58, 93, 102]. They usually formalize the input user sequence as a soft prompt, and map the label to the answer word by a designed verbalizer. For example, Zhang and Wang [102] transformed the news recommendation task into a cloze-style mask-prediction task under the prompt learning paradigm.

Generative language models such as ChatGPT exhibit superior capabilities in natural language generation. Many recent research efforts have been made to leverage the impressive power of generative language models by translating the recommendation task into natural language. Specifically, the few-shot and zero-shot abilities are first explored in the recommendation task. One line of research [12, 46, 66, 71, 89] tested the recommendation performance with a few training samples. For example, Liu et al. [46] and Dai et al. [12] formulated the recommendation prompts for different tasks and evaluated the performance using ChatGPT. Some works also tried the use of in-context learning by injecting a few demonstration examples into the prompts to trigger the few-shot abilities without the update of parameters. Another line of works [2, 3, 22, 33, 35, 42, 45] devoted to better utilizing the training samples by introducing the prompt tuning. For example, Geng et al. [22] fine-tuned the T5 model [63] with five different instruction types to support diverse downstream recommendation tasks in a unified framework. Bao et al. [3; 2] proposed a two-stage fine-tuning framework to incrementally adapt LLMs to recommendation tasks. Lin et al. [45] employed multi-facet token identifiers and specialized data structures to encourage valid recommendations by LLMs. However, fine-tuning LLMs often requires sophisticated training paradigms and can increase the efficiency burden on the recommendation model. Recently, some works Ren et al. [64] and Wei et al. [86] have also explored using LLMs to generate augmented data for enhancing recommendations. For example, Wei et al. [86] generated augmented implicit feedback along with user and item profiles to improve recommendation performance. Nevertheless, these approaches do not effectively incorporate collaborative knowledge, and the quality of the augmented items remains difficult to assess.

In this article, we leverage LLMs as pseudo-cross-domain item predictors to retrieve the latent cross-domain items in the single-domain sequences. We first design a candidate-free cross-domain interaction generation task for LLMs and propose corresponding solutions to supplement the collaborative knowledge and assess the relevance of generated interactions.

3 Problem Definition

In our study, we address the CDSR problem, particularly focusing on scenarios with both overlapped and non-overlapped users. Specifically, we denote the two domains as A and B. Let $v^A \in \mathcal{V}^A$ and $v^B \in \mathcal{V}^B$ represent the item in domain A and B, respectively. Each item is equipped with textual attributes, such as title and description. Note that $\mathcal{V}^A$ and $\mathcal{V}^B$ denote the entire item set in domain A and B, respectively. In our problem, the users in domains A and B are not completely the same. Therefore, we use $\mathcal{U}^O$, $\mathcal{U}^A$, and $\mathcal{U}^B$ to denote the sets of overlapped users and non-overlapped users in domains A and domain B respectively. We use $\mathcal{S}^O$, $\mathcal{S}^A$, and $\mathcal{S}^B$ to represent the sets of historical interactions of these users. For a non-overlapped user $u \in \mathcal{U}^A$ in domain A, the user's historical interaction sequence $S^A \in \mathcal{S}^A$ in domain A are represented as $S^A = \left[v_1^A, v_2^A, \ldots, v_\tau^A\right]$ where τ is the length of the sequence (Similarly for S^B). For an overlapped user $u \in \mathcal{U}^O$, we merge the sequences S^A and S^B in two domains in chronological order, and denote the mixed cross-domain interaction sequence as $S^O \in \mathcal{S}^O$. We summarize the frequently used notations in Table 1.

Formally, given the sets of historical interactions $\mathcal{S}^O$, $\mathcal{S}^A$ and $\mathcal{S}^B$, we aim to learn a recommender $\mathcal{F}$. Then, given the observed interactions S_t^O at time t of a new coming overlapped user u^*, along

Table 1. Mathematical Notations

Notation	Description
$\mathcal{V}^X$	The sets of items in domain X
$\mathcal{U}^O, \mathcal{U}^X$	The sets of overlapped users, and non-overlapped users in domain X
$\mathcal{S}^O, \mathcal{S}^X$	The sets of historical interactions of overlapped users, and non-overlapped users in domain X
$g^X, \boldsymbol{g}^X, \widehat{\boldsymbol{g}}^X$	The generated interactions in domain X by LLM, its textual and inferred collaborative embedding
$\boldsymbol{M}^X, \boldsymbol{M}^{X*}$	The learnable projection matrix and its trained version
$\boldsymbol{w}$	The parameter in the relevance weighting layer
$\widehat{\mathcal{V}}^X, \widehat{S}^{A\leftarrow B}, \widehat{S}^{B\leftarrow A}$	The pseudo cross-domain items and sequences
$\widehat{\mathcal{S}}^O$	The set of pseudo cross-domain sequences of all non-overlapped users

X refers to either domain A or domain B.

with the domain to which the next item belongs, the goal is to predict the next item v_{t+1} by selecting the one with the highest probability:

$$v_{t+1} = \begin{cases} \arg\max_{v_i^A \in \mathcal{V}^A} P(v_i^A | S_t^O), & \text{if } v_{t+1} \in \mathcal{V}^A, \\ \arg\max_{v_i^B \in \mathcal{V}^B} P(v_i^B | S_t^O), & \text{if } v_{t+1} \in \mathcal{V}^B, \end{cases} \tag{1}$$

where $P(v_i^A|S^O) \in [0,1]^{|\mathcal{V}^A|}$ and $P(v_i^B|S^O) \in [0,1]^{|\mathcal{V}^B|}$ are the probabilities of items in domain A and B, as predicted by $\mathcal{F}$.

4 Methodology

In this section, we present the proposed LLMCDSR for CDSR. We first introduce the overall framework and then describe the details of each component.

4.1 Framework Overview

Figure 2 presents the overview of our proposed framework LLMCDSR. In specific, we start with the *Candidate-Free Cross-Domain Interaction Generation* (Section 4.2), where LLMs are tasked with generating cross-domain interaction based on the single-domain user interactions without the candidate selection procedure. Then, we introduce a *Collaborative-Textual Contrastive Pre-Training* strategy (Section 4.3), which learns the textual projection matrices for supplementing the collaborative knowledge into the textual features of LLM generations. Building upon this, we present an *RMRN* (Section 4.4) which features a relevance weighting layer, designed to adaptively evaluate the relevance and quality of the LLM-generated samples in relation to the downstream task. A recall operator is then employed to retrieve pseudo items in the dataset based on the weighted generation samples. The enhanced single-domain sequences are incorporated to optimize the parameters of the relevance weighting layer alongside the downstream cross-domain recommender in a meta-learning manner.

4.2 Candidate-Free Cross-Domain Interaction Generation

In the more general scenario of the CDSR problem, the presence of the overlapped users is often limited, and a significant number of non-overlapped users, who interact within only one specific domain, abound. The challenge lies in leveraging the information from these single-domain interactions to enhance CDSR, as the lack of observations in other domains hampers cross-domain knowledge transfer. With LLMs encapsulating a wide-ranging knowledge base, they could also potentially include implicit cross-domain information. For example, text such as related books for a movie can be usually found on the Internet. Therefore, we leverage LLMs to infer the latent missing

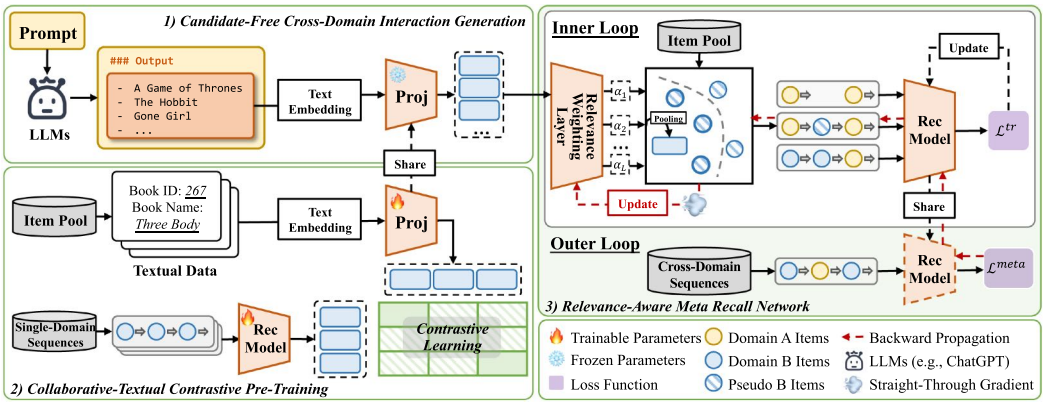


Fig. 2. The overview of our proposed LLMCDSR framework.

cross-domain items in single-domain sequences, thereby breaking the knowledge barriers between domains in these sequences.

Recent works utilizing LLMs for recommendation tasks often face inherent limitations in handling extensive candidate sets and complex collaborative information. Typically, these approaches [3, 37] provide LLMs with a pre-selected list of candidate items, from which the LLMs choose recommendations. Such a method can result in biased candidate selection and heavily depends on the quality of the sampling strategy. Instead of taking LLMs as the recommenders, we first introduce a novel paradigm in this subsection where we harness LLM's few-shot abilities to generate cross-domain interactions in a candidate-free manner without the candidate selection procedure.

Specifically, we aim to build a prompt template $T(\cdot,\cdot)$ that guides LLMs to generate cross-domain interactions. The prompt begins with an instruction detailing the primary task, *i.e.,* to generate interactions for one domain based on interactions in another domain. In line with the concept of in-context learning [15, 103], the prompt also includes several contextual exemplars to aid the generation process. These examples are selected from the original cross-domain interactions of overlapped users. Finally, the input single-domain sequence is fed into the LLMs. The overview of the prompt template is shown in Figure 3, and the complete version can be found in Appendix A.1. Formally, given a non-overlapped user in domain A with single-domain interactions S^A, the candidate-free cross-domain interaction generation is conducted as follows:

$$\left\{g_1^B, g_2^B, \ldots, g_L^B\right\} = \mathsf{LLM}\left(T\left(S^A, \left\{(S^O)_i\right\}_{i=1}^m\right)\right), (S^O)_i \sim \mathcal{S}^O, \tag{2}$$

where $(S^O)_i$ is the in-context exemplar sampled from the cross-domain interaction set, and $\left\{g_1^B, g_2^B, \ldots, g_L^B\right\}$ is the text-like generated interactions by LLMs and L is the size of the items. Note that the generation of interactions $\left\{g_1^A, g_2^A, \ldots, g_L^A\right\}$ for non-overlapped users in domain B can be similarly obtained with Equation (2).

Since our framework only requires the inference ability of LLMs for generating cross-domain interactions, it offers strong generalizability and compatibility with other methods aimed at enhancing the quality of LLM-generated content. For example, fine-tuned LLMs on specific recommendation datasets can be seamlessly integrated into our framework to generate higher-quality interactions. We will leave the exploration of these improvements as part of our future endeavors.

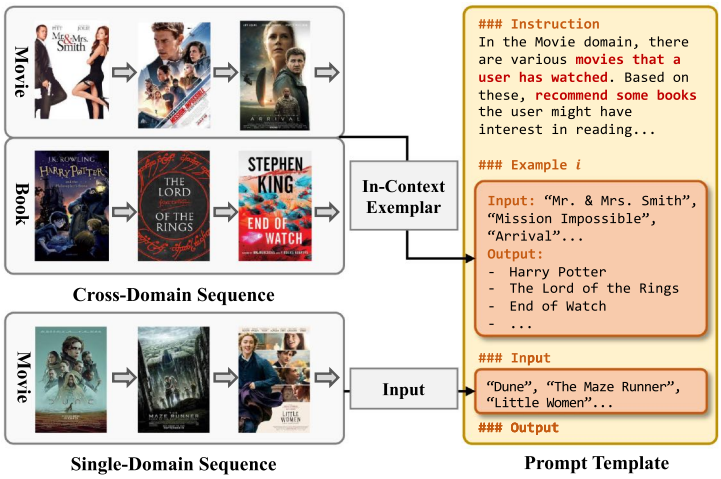


Fig. 3. The demonstration of the prompt for the candidate-free cross-domain interaction generation.

4.3 Collaborative-Textual Contrastive Pre-Training

The LLM generations of cross-domain interactions are yielded based on purely textual descriptions of a single user's historical interactions, assisted by several in-context cross-domain samples. Consequently, such solely candidate-free generations may lead LLMs to overlook collaborative information, which is crucial for estimating the behaviors. Fortunately, the textual attributes of items within the dataset present an opportunity. By utilizing item-text pairs from the dataset, we can learn to infuse the collaborative knowledge of the dataset into these generated samples. To achieve this, we propose a Collaborative-Textual Contrastive Pre-Training strategy, where we learn a textual projection matrix through a contrastive learning objective. This pre-trained projection matrix is then employed to obtain the inferred collaborative representations, which are informed by and attuned to the collaborative signals present in the dataset, thereby enhancing their relevance and utility in the recommendation process.

4.3.1 Sequential Collaborative Representation Learning. Specifically, we first define a sequential recommender consisting of two groups of parameters: a randomly initialized embedding matrix $\boldsymbol{P}^X \in \mathbb{R}^{|\mathcal{V}^X| \times d}$ for the items where d is the dimension of the embeddings, and a sequential encoder $\mathsf{SeqEnc}^X_{pt}(\cdot)$ to encode the interaction sequence, which can be instantiated as advanced neural networks such as GNNs [88], GRUs [28], and self-attention networks [36]. The embedding matrix $\boldsymbol{P}^X$ and the parameters in $\mathsf{SeqEnc}^X_{pt}(\cdot)$ are collectively denoted as Θ^X_{pt}, *i.e.*, $\Theta^X_{pt} = \left\{\boldsymbol{P}^X, \mathsf{SeqEnc}^X_{pt}(\cdot)\right\}$. Here, X refers to either domain A or domain B, indicating that identical initializations and operations are applied to both domains. For the sake of clarity and simplicity in our explanation, we will continue to use this notation throughout this article.

Given a non-overlapped user's the single-domain sequence $S^X_t = \left[v^X_1, v^X_2, \ldots, v^X_t\right]$ of domain X, it is encoded into a low-dimensional representation $\boldsymbol{h}^X_t$ as follows:

$$\boldsymbol{h}^X_t = \mathsf{SeqEnc}^X_{pt}\left(S^X_t\right). \tag{3}$$

To fully take advantage of the single-domain signals in the overall interactions, we also extract the single-domain sequence $S^{O(X)}$ from the interactions of overlapped users $S^O \in \mathcal{S}^O$, and denote the set of them as $\mathcal{S}^{O(X)}$. Using the sequence sets $\mathcal{S}^X$ and $\mathcal{S}^{O(X)}$, we adopt the binary cross entropy

loss as the objective loss function to pre-train the sequential recommender:

$$\mathcal{L}_X^{ptr} = \sum_{S \in \mathcal{S}^{\{X,O(X)\}}} \ell_S^{ptr,X} = - \sum_{S \in \mathcal{S}^{\{X,O(X)\}}} \sum_{t=1}^{\tau-1} \left[\log P_{pt}^X(v_i^X | S_t) + \log \left(1 - P_{pt}^X(v_j^X | S_t) \right) \right], \tag{4}$$

where $v_i^X = v_{t+1}^X$ is the ground truth item interacted at time $t+1$, $v_j^X \notin S$ is a randomly sampled negative item, and the probability P_{pt}^X is calculated by the inner product as:

$$P_{pt}^X(v_i^X | S_t) = \sigma \left(\left\langle \boldsymbol{h}_t^X, \boldsymbol{p}_i^X \right\rangle \right), \tag{5}$$

where $\langle \cdot \rangle$ is the inner product, σ is the sigmoid function, and $\boldsymbol{p}_i^X$ is the ith row of $\boldsymbol{P}^X$ (i.e., the embedding for item v_i^X).

4.3.2 Learnable Textual Projection. We make use of the textual information of the items in the dataset as anchors to learn a projection matrix. Specifically, given an item $v_i^X \in \mathcal{V}^X$ with its textual information c_i^X, we hire a text embedding model $\texttt{TxtEmb}(\cdot)$ to encode the textual description into dense embedding $\boldsymbol{c}_i^X \in \mathbb{R}^{d_c}$ (d_c is the dimension of the text embedding):

$$\boldsymbol{c}_i^X = \texttt{TxtEmb}(c_i^X). \tag{6}$$

After the textual embedding of item v_i^X is obtained, we employ a learnable projection matrix $\boldsymbol{M}^X \in \mathbb{R}^{d_c \times d}$ to obtain the inferred collaborative embedding from the textual embedding as follows:

$$\widetilde{\boldsymbol{c}}_i^X = \left(\boldsymbol{M}^X \right)^{\mathrm{T}} \boldsymbol{c}_i^X, \tag{7}$$

where $\widetilde{\boldsymbol{c}}_i^X \in \mathbb{R}^d$ is the inferred collaborative embedding of v_i^X. The embedding, derived from item textual features, incorporates collaborative information through the textual projection matrix.

4.3.3 Contrastive Alignment. Inspired by recent advancement of contrastive learning in self-supervised learning [8, 40, 61], we propose to align the item embedding and the inferred collaborative embedding from textual features by predicting whether an item-text pair comes from the same item. Given one batch with n items, we have the following contrastive loss function:

$$\ell_i^{ptc,X} = -\log \frac{e^{\langle \boldsymbol{p}_i^X, \widetilde{\boldsymbol{c}}_i^X \rangle / \gamma}}{\sum_{j=1}^n e^{\left\langle \boldsymbol{p}_i^X, \widetilde{\boldsymbol{c}}_j^X \right\rangle / \gamma}} - \log \frac{e^{\langle \widetilde{\boldsymbol{c}}_i^X, \boldsymbol{p}_i^X \rangle / \gamma}}{\sum_{j=1}^n e^{\left\langle \widetilde{\boldsymbol{c}}_i^X, \boldsymbol{p}_j^X \right\rangle / \gamma}}, \tag{8}$$

where γ is the scale parameter. Finally, we combine the recommendation loss and the contrastive loss, and jointly pre-train the parameters Θ_{pt}^X of the recommender and the projection matrix $\boldsymbol{M}^X$:

$$\mathcal{L}_X^{pt}(\Theta_{pt}^X, \boldsymbol{M}^X) = \lambda_c \sum_{i=1}^{|\mathcal{V}^X|} \ell_i^{ptc,X} + \mathcal{L}_X^{ptr}, \tag{9}$$

where λ_c is a hyper-parameter. And we denote the trained projection matrix as $\boldsymbol{M}^{X*}$.

4.4 Relevance-Aware Meta Recall Network

While the cross-domain interactions generated by LLMs offer a valuable resource, their direct application in specific downstream recommendation tasks is not straightforward. Firstly, the items recommended in a candidate-free manner by LLMs do not correspond to the items in the dataset. They are typically described using a few natural language words. Additionally, due to the acknowledged hallucination problem [99] in LLMs, the quality of their outputs can be inconsistent and difficult to evaluate. In other words, the LLM-generated content is potentially noisy and unstable, and it's challenging to determine which outputs are genuinely beneficial for

the specific recommendation task at hand. To meet these, we propose an RMRN to automatically weigh the relevance of generated samples, and then produce relevant pseudo-cross-domain items from the dataset. RMRN starts with a relevance weighting layer which assesses the significance of the generated samples. Then, it recalls cross-domain items from the dataset, forming pseudo-cross-domain sequences to facilitate CDSR. Subsequently, a bi-level meta-optimization strategy, incorporating relevance awareness, is employed to train the parameters of the relevance weighting layer and the downstream recommender.

4.4.1 Relevance Weighting Layer. With the interaction samples generated by LLMs in Equation (2), we also use the text embedding model to encode them into representations as:

$$\left\{\boldsymbol{g}_1^X, \boldsymbol{g}_2^X, \ldots, \boldsymbol{g}_L^X\right\} = \texttt{TxtEmb}\left(\left\{g_1^X, g_2^X, \ldots, g_L^X\right\}\right). \tag{10}$$

To better align the LLM textual generation space and the item space, we fetch the pre-trained textual projection matrix $\boldsymbol{M}^{X*}$ in Section 4.3 to obtain the inferred collaborative embeddings from the above textual embeddings:

$$\widetilde{\boldsymbol{g}}_i^X = \left(\boldsymbol{M}^{X*}\right)^{\mathrm{T}} \boldsymbol{g}_i^X, \quad i = 1, 2, \ldots, L. \tag{11}$$

As introduced in Section 4.3, the inferred collaborative embedding $\widetilde{\boldsymbol{g}}_i^X$ is more informative since the collaborative knowledge is incorporated through $\boldsymbol{M}^{X*}$. Afterwards, as different generation samples may have different significance w.r.t. the relevance for the downstream recommendation task, we propose a relevance weighting layer to automatically weigh each generation sample as follows:

$$\begin{aligned} \pi_i &= \boldsymbol{w}^{\mathrm{T}} \widetilde{\boldsymbol{g}}_i^X, \\ \alpha_i &= \frac{e^{\pi_i}}{\sum_{j=1}^{L} e^{\pi_j}}, \quad i = 1, 2, \ldots, L. \end{aligned} \tag{12}$$

where $\boldsymbol{w} \in \mathbb{R}^d$ is a trainable parameter. After obtaining the weight scored by the relevance weighting layer, we can perform a weighted pooling operation to selectively aggregate the generation samples into a consolidated representation as:

$$\boldsymbol{r}^X = \frac{1}{L} \sum_{i=1}^{L} \alpha_i \widetilde{\boldsymbol{g}}_i^X, \tag{13}$$

where $\boldsymbol{r}^X \in \mathbb{R}^d$ is coined as the generation anchor representation. To adapt the LLM generation samples to the dataset items, as the item embeddings and the generation anchor representations are actually in the same collaborative feature space, we employ a recall function to retrieve K items from the dataset based on the representation similarity as:

$$\left\{\widehat{v}_1^X, \widehat{v}_2^X, \ldots, \widehat{v}_K^X\right\} = \underset{i=1,2,\ldots,|\mathcal{V}^X|}{\texttt{Top-K}} \left(\left\langle \boldsymbol{p}_i^X, \boldsymbol{r}^X \right\rangle\right), \tag{14}$$

where $\texttt{Top-K}(\cdot)$ is designed to identify the K nearest indices according to a specific metric. Using this function, we can select K items in domain B whose embeddings are the K most similar to the anchor representation.

For each non-overlapped user, we denote the set of these recalled items as $\widehat{\mathcal{V}}^X = \left\{\widehat{v}_1^X, \widehat{v}_2^X, \ldots, \widehat{v}_K^X\right\} \subset \mathcal{V}^X$. With these items in the dataset retrieved based on the LLM generation, they can be seen as the pseudo cross-domain items to be incorporated into the downstream CDSR task.

4.4.2 Relevance-Aware Meta Optimization. To effectively utilize the cross-domain knowledge derived from LLMs, we first integrate these variable pseudo-cross-domain items into the training dataset. As the original cross-domain sequences from overlapped users offer a valuable reference for ideal cross-domain behaviors, we regard them as the meta samples and introduce a bi-level optimization in a meta-learning manner to update the parameters of both the relevance weighting layer and the downstream recommender. Such an optimization ensures that items of higher quality are preferentially recalled, thereby augmenting the CDSR with these pseudo-cross-domain sequences as training samples.

Pseudo Cross-Domain Sequence Construction. Recall that for each single-domain sequence $S^A \in \mathcal{S}^A, S^B \in \mathcal{S}^B$, we have obtained the pseudo cross-domain items $\widehat{\mathcal{V}}^B \subset \mathcal{V}^B, \widehat{\mathcal{V}}^A \subset \mathcal{V}^A$ based on the candidate-free cross-domain interaction generations from LLMs, respectively. As [51, 90] suggests, the order of users' interactions in the real world is often influenced by various unobservable external factors, leading to a natural flexibility in interaction sequences. Therefore, we randomly insert these pseudo items into the single-domain sequence to enhance the robustness especially when the sequence order may not be as strictly defined. Taking S^A as an example, the single-domain sequence is augmented into the pseudo-cross-domain sequence $\widehat{S}^{A \leftarrow B}$ as:

$$\widehat{S}^{A \leftarrow B} = \texttt{RandomInsert}(S^A, \widehat{\mathcal{V}}^B). \tag{15}$$

The set of the pseudo-cross-domain sequences $\widehat{S}^{A \leftarrow B}$ and $\widehat{S}^{B \leftarrow A}$ of all non-overlapped users in the training set is denoted as $\widehat{\mathcal{S}}^O$. As these sequences include items from both domains A and B, they can act as the bridge to assist in the knowledge alignment and transfer across domains.

Meta Learning Objectives. Similarly, with Section 4.3.1, we define another unified sequential recommender that takes the mixed sequences containing items from both domain A and B as input, whose parameters are denoted as $\Theta_{rec} = \left\{\boldsymbol{V}^A, \boldsymbol{V}^B, \texttt{SeqEnc}_{rec}(\cdot)\right\}$, where $\boldsymbol{V}^A \in \mathbb{R}^{|\mathcal{V}^A| \times d}$ and $\boldsymbol{V}^B \in \mathbb{R}^{|\mathcal{V}^B| \times d}$ are another two embedding matrices for items in domain A and B respectively. Also, we define the loss function for the unified sequential recommender Θ_{rec} as:

$$\begin{aligned}
\ell_S^{rec} &= -\sum_{t=1}^{\tau-1} \mathbb{1}(v_{t+1}) [\log P_{rec}(v_i|S_t) + \log\left(1 - P_{rec}(v_j|S_t)\right)], \\
P_{rec}(v_i|S_t) &= \sigma\left(\left\langle \boldsymbol{h}'_t, \boldsymbol{v}_i \right\rangle\right), \quad \boldsymbol{h}'_t = \texttt{SeqEnc}_{rec}(S_t), \\
\mathbb{1}(v_{t+1}) &= \begin{cases} 0 & \text{if } v_{t+1} \text{ is a pseudo item,} \\ 1 & \text{otherwise,} \end{cases}
\end{aligned} \tag{16}$$

where $\mathbb{1}(v_{t+1})$ is an indicating function making sure that the training objective is not disturbed by the recalled pseudo items. Since now we have both the original interaction data $\mathcal{S}^A, \mathcal{S}^B, \mathcal{S}^O$ and the pseudo cross-domain interaction data $\widehat{\mathcal{S}}^O$, the optimal parameters Θ^*_{rec} can be subsequently calculated as follows:

$$\Theta^*_{rec}(\boldsymbol{w}) = \arg\min_{\Theta_{rec}} \mathcal{L}^{tr}(\Theta_{rec}; \boldsymbol{w}) \triangleq \arg\min_{\Theta_{rec}} \sum_{S \in \mathcal{S}^{\{A,B,O\}} \cup \widehat{\mathcal{S}}^O} \ell_S^{rec}. \tag{17}$$

Note that $\mathcal{L}^{tr}(\Theta_{rec}; \boldsymbol{w})$ means that the loss is calculated under the influence of the relevance weighting layer (i.e., $\boldsymbol{w}$). This is because the items recalled in the pseudo-cross-domain sequences $\widehat{\mathcal{S}}^O$ are generated through the generation anchor representation in Equation (13), which is parameterized by $\boldsymbol{w}$. Consequently, the parameters of the unified recommender are functionally dependent on $\boldsymbol{w}$, that is, $\Theta_{rec} = \Theta_{rec}(\boldsymbol{w})$. Moreover, to ensure that the relevance weighting layer adaptively weighs the noisy generation samples, which could be detrimental to the downstream recommendation task, we adopt the idea of meta-learning [20, 21, 68] to optimize the parameter $\boldsymbol{w}$ of the relevance

weighting layer. Specifically, as the data from original cross-domain sequences of overlapped users provide a gold standard for how actual cross-domain sequences should manifest, they serve as meta samples to offer meta-knowledge of the ground-truth cross-domain interaction distribution. The parameter $\boldsymbol{w}$ is optimized by the defined meta loss as follows:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \mathcal{L}^{meta}(\Theta^*_{rec}(\boldsymbol{w})) \triangleq \arg\min_{\boldsymbol{w}} \sum_{S \in \mathcal{S}^O} \ell^{rec}_S. \tag{18}$$

Bi-Level Optimization. Since the searching for optimal Θ^*_{rec} and $\boldsymbol{w}^*$ requires two nested loops of optimization [20], we employ an online strategy to update Θ^*_{rec} and $\boldsymbol{w}^*$ with a bi-level optimization manner. As shown in Figure 2, the optimization procedure is conducted as:

— *Assumed Update of Θ_{rec} (Inner Loop).* An assumed update of Θ_{rec} is performed with fixing the parameter $\boldsymbol{w}$:

$$\widehat{\Theta}_{rec} = \Theta_{rec} - \eta_1 \nabla_{\Theta_{rec}} \mathcal{L}^{tr}(\Theta_{rec}; \boldsymbol{w}), \tag{19}$$

where η_1 is the learning rate of the assumed update.

— *Meta Update of $\boldsymbol{w}$ (Outer Loop).* As shown by the red dashed arrow in Figure 2, the update of $\boldsymbol{w}$ can be instructed by the gradient of the meta samples (i.e., the original cross-domain sequences) on the assumed-updated model:

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta_2 \nabla_{\boldsymbol{w}} \mathcal{L}^{meta}(\widehat{\Theta}_{rec}(\boldsymbol{w})), \tag{20}$$

where η_2 is the learning rate of the meta update.

To enhance the flexibility of our framework, we also set the meta update interval Δu as a hyper-parameter to control the update frequency of the parameter $\boldsymbol{w}$. Notably, as we can find that the backward propagation of the meta loss is cut off due to the non-differentiable recall operator (Equation (14)), it is non-trivial to update the parameter $\boldsymbol{w}$ using Equation (20). Therefore, we employ the straight-through estimation [9, 76] to proximate the gradient of $\boldsymbol{w}$, which is derived as:

$$\nabla_{\boldsymbol{w}} \mathcal{L}^{meta}(\widehat{\Theta}_{rec}(\boldsymbol{w})) \tag{21a}$$

$$= \frac{1}{|\mathcal{S}^O|} \sum_{S \in \mathcal{S}^O} \nabla_{\boldsymbol{w}} \ell^{rec}_S \tag{21b}$$

$$= -\frac{\eta_1}{|\mathcal{S}^O|} \sum_{S \in \mathcal{S}^O} \frac{\partial \ell^{rec}_S}{\partial \widehat{\Theta}_{rec}(\boldsymbol{w})} \left(\frac{1}{|\widehat{\mathcal{S}}^O|} \sum_{S' \in \widehat{\mathcal{S}}^O} \frac{\partial \widehat{\Theta}_{rec}(\boldsymbol{w})}{\partial \nabla_{\Theta_{rec}} \ell^{rec}_{S'}} \frac{\partial \nabla_{\Theta} \ell^{rec}_{S'}}{\partial \widehat{\boldsymbol{V}}_{S'}} \frac{\partial \widehat{\boldsymbol{V}}_{S'}}{\partial \boldsymbol{r}} \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{w}} \right) \tag{21c}$$

$$\approx -\frac{\eta_1}{|\mathcal{S}^O||\widehat{\mathcal{S}}^O|} \sum_{S \in \mathcal{S}^O} \frac{\partial \ell^{rec}_S}{\partial \widehat{\Theta}_{rec}(\boldsymbol{w})} \sum_{S' \in \widehat{\mathcal{S}}^O} \left(\frac{\partial \widehat{\Theta}_{rec}(\boldsymbol{w})}{\partial \nabla_{\Theta} \ell^{rec}_{S'}} \frac{1}{K} \left(\sum_{i=1}^{K} \frac{\partial \nabla_{\Theta} \ell^{rec}_{S'}}{\partial \widehat{\boldsymbol{v}}_i} \right) \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{w}} \right), \tag{21d}$$

where $\widehat{\boldsymbol{V}}_{S'}$ denotes the set of embeddings for pseudo items in sequence S', with $\widehat{\boldsymbol{v}}_i$ being their respective embeddings. Due to the non-differentiable nature of the recall operator, the derivative $\frac{\partial \widehat{\boldsymbol{V}}_{S'}}{\partial \boldsymbol{r}}$ is not valid in Equation (21c). To address this, we approximate it to 1 and average the gradients over $\widehat{\boldsymbol{V}}_{S'}$ for the sake of dimensional consistency. In practice, this means that during the backpropagation, we transfer the average gradients from $\widehat{\boldsymbol{V}}_{S'}$ directly to $\boldsymbol{r}$. After the meta update of parameter $\boldsymbol{w}$, we can finally perform the actual update of the recommender parameters Θ_{rec} as:

$$\Theta_{rec} \leftarrow \Theta_{rec} - \eta_1 \nabla_{\Theta_{rec}} \mathcal{L}^{tr}(\Theta_{rec}; \boldsymbol{w}). \tag{22}$$

4.5 Model Learning

Generally, in our framework LLMCDSR, we first leverage LLMs to generate cross-domain interactions without a candidate selection procedure. Then, we pre-train the textual projection matrix for each domain. Utilizing these trained matrices, we infuse the text embeddings of the generated samples with collaborative knowledge, where the inferred collaborative representations are obtained. Subsequently, the significance of these generated samples is weighted based on their relevance and quality concerning the downstream recommendation task. This is followed by the implementation of a recall operator, which identifies and retrieves relevant items from the dataset. The recalled items are then utilized to construct pseudo-cross-domain sequences, enhancing knowledge alignment transfer across domains for effective CDSR. Both the relevance weighting layer and the ultimate cross-domain recommender are optimized in a bi-level meta-learning manner with the guidance of the original cross-domain interactions. The learning procedure of our algorithm is shown in Algorithm 1.

5 Experiments

In this section, we conduct experiments on two widely used public datasets to evaluate the effectiveness of the proposed LLMCDSR by answering the following **Research Questions (RQs)**:

- *RQ1*: How does LLMCDSR perform compared with the baseline methods in terms of NDCG and Hit Ratio? (See Section 5.3)
- *RQ2*: How much do different components in LLMCDSR contribute to the recommendation performance? (See Section 5.4)
- *RQ3*: How do the hyper-parameter settings in LLMCDSR, such as the number of recalled items and the meta-update interval, affect the recommendation performance? (See Section 5.5)
- *RQ4*: How do LLMCDSR and competitive baselines perform across different levels of user overlap? (See Section 5.6)
- *RQ5*: How does LLMCDSR perform when equipped with different LLM backbones? (See Section 5.7)
- *RQ6*: How does the efficiency of LLMCDSR compare to the baseline methods in terms of training and inference time? (See Section 5.8)
- *RQ7*: How does LLMCDSR effectively retrieve pseudo-cross-domain items to facilitate the CDSR? (See Section 5.9)

5.1 Datasets

Following previous works, we built two CDSR datasets[1] based on the publicly available Amazon[2] datasets. Specifically, to examine the effectiveness of our framework across different levels of domain correlations, we selected four domains to generate two cross-domain pairs for evaluation, i.e., *Movie-Book* and *Food-Kitchen*. The Movie-Book represents domains with closer relations, whereas the Food-Kitchen demonstrates domains with less direct relevance to each other. To enable the LLMs for open cross-domain interaction generation based on textual description, we first removed the items that lacked textual data. For each cross-domain pair, to assure data quality, we filtered out users and items of which the number of interactions is fewer than 10 for Movie-Book dataset and 5 for Food-Kitchen dataset. Note that as we studied the CDSR problem featuring both overlapped and non-overlapped users in this article, we not only preserved users who have interactions in both domains, but also those who only have interactions in a single domain. Moreover, to make the interactions satisfy the sequential nature, we ordered the interactions by time for each user, then

[1]To further assess the generalizability of our framework, we perform additional experiments on a supplementary dataset, detailed in Appendix A.2.

[2]https://jmcauley.ucsd.edu/data/amazon/index_2014.html

Algorithm 1: Training Procedure of LLMCDSR

Input : LLMs $\texttt{LLM}(\cdot)$, text embedding model $\texttt{TxtEmb}(\cdot)$ historical interaction sets $\mathcal{S}^{\{O,A,B\}}$, item sets $\mathcal{V}^{A,B}$

Output : The cross-domain sequential recommender $\mathcal{F}$ with parameters Θ_{rec}

for *each sequence* S *in* $\mathcal{S}^A \cup \mathcal{S}^B$ **do**
  Get cross-domain interaction generations $\{g_i\}_{i=1}^L$ from LLMs following Equation (2);
end
Initialize pre-train parameters: recommenders $\Theta_{pt}^{\{A,B\}}$, textual projection matrices $\boldsymbol{M}^{\{A,B\}}$;
for *domain* X *in* $\{A, B\}$ **do**
  while *not converged* **do**
    Calculate the pre-train loss by Equation (9);
    Update parameters Θ_{pt}^X and $\boldsymbol{M}^X$;
  end
  Get the pre-trained textual projection matrix $\boldsymbol{M}^{X*}$;
end
Initialize parameters: downstream cross-domain recommender Θ_{rec}, relevance weighting layer parameter $\boldsymbol{w}$;
for *domain* X *in* $\{A, B\}$ **do**
  for *each sequence* S *in* $\mathcal{S}^X$ **do**
    Get the textual embeddings $\left\{\boldsymbol{g}_i^X\right\}_{i=1}^L$ of the generation samples $\left\{g_i^X\right\}_{i=1}^L$ from LLMs;
    Get the inferred collaborative embeddings $\left\{\widetilde{\boldsymbol{g}}_i^X\right\}_{i=1}^L$ by Equation (11);
  end
end
while *not converged* **do**
  $\widehat{\mathcal{S}}^O = \emptyset$;
  for *domain* X *in* $\{A, B\}$ **do**
    for *each sequence* S *in* $\mathcal{S}^X$ **do**
      Get the generation anchor representation $\boldsymbol{r}^X$ through the relevance weighting layer by Equations (12) and (13);
      Recall K items $\widehat{\mathcal{V}}_u^X$ in the dataset using $\boldsymbol{r}^X$ by Equation (14);
      Get the pseudo cross-domain sequence $\widehat{S}$ by Equation (15);
      $\widehat{\mathcal{S}}^O = \widehat{\mathcal{S}}^O \cup \{\widehat{S}\}$;
    end
  end
  if *updating meta parameter* $\boldsymbol{w}$ **then**
    Assumed update of Θ_{rec} and get $\widehat{\Theta}_{rec}$ by Equation (19);
    Meta update of $\boldsymbol{w}$ by Equation (20);
  end
  Actual update of Θ_{rec} by Equation (22);
end
return $\mathcal{F}_{\Theta_{rec}}$

Table 2. Dataset Statistics

Dataset		Movie-Book		Food-Kitchen	
		Movie	Book	Food	Kitchen
# Users	Non-overlapped	6,819	50,242	5,041	35,889
	Overlapped	2,666		5,781	
# Interaction records	Single-domain	12,185	124,889	5,472	38,191
	Cross-domain	6,652		7,501	
# Items		12,875	93,860	8,661	27,637
Average sequence length		9.01		8.22	
Density		0.0099%		0.0226%	

split the sequences from each user into several small sequences with each sequence containing interactions within a period, i.e., a month for the Movie-Book dataset, and a year for the Food-Kitchen dataset. For each cross-domain sequence, it should meet the requirement containing at least 2 items from each domain, while for each single-domain sequence, the minimum number of items is 5. In terms of evaluation, we randomly split the sequences into a training set, a validation set, and a test set following the proportions: 80%/10%/10%, while guaranteeing that the sequences of the user in the validation and test set happen after the user's corresponding sequences in the training set to avoid the information leak issue [5]. Finally, we used the last interacted item in each sequence as the ground truth item to investigate the performance, respectively. The statistics of the two datasets are shown in Table 2.

5.2 Experimental Setups

5.2.1 Baselines. To validate the effectiveness of our proposed model, we compare it with three categories of baselines: (1) single-domain SR methods; (2) CDSR methods.

(1) Single-domain SR methods:

— *FPMC* [65] combines personalized transition graphs and MC and proposes a factorized personalized Markov chain model to decompose the transition cube in the sequential data.
— *Caser* [74] transforms sequences of items into an "image" in time and latent spaces and uses convolutional filters to learn sequential patterns as local features of this image.
— *GRU4Rec* [28] uses RNNs to model the sequential interactions for the session-based recommendation.
— *SRGNN* [88] uses GNNs to capture complex item transitions of session sequences modeled as graph-structured data and employs an attention network to represent each session through a combination of global preference and current interest.
— *FEARec* [17] learns both low- and high-frequency information in the frequency domain with a ramp structure and incorporates an auto-correlation-based attention mechanism to capture periodic characteristics.
— *SASRec* [36] utilizes a self-attention-based sequential model to balance capturing long-term semantics with making predictions based on fewer actions.

(2) CDSR methods:

— *TPUF* [14] proposes a feature mapping-and-aggregation framework, which transfers pre-trained latent user features from the source domain to address the privacy concerns.

— π **-Net** [56] consists of two gating components for filtering and transferring information between domains based on overlapped users, and is the pioneering work for CDSR.
— *C2DSR* [5] utilizes GNNs and self-attention networks to encode intra- and inter-sequential relationships, and presents a contrastive infomax objective to strengthen the correlations between single- and cross-domain representations.
— ***MGCL*** [92] utilizes contrastive learning in two views of intra-domain item representation and inter-domain user preference to jointly learn the dynamic sequential information and complementary preference information.

5.2.2 Evaluation Protocol. As previous works suggest [5, 36], we adopt the leave-one-out method to calculate the recommendation performance. For each validation and test instance, we treat 999 negative items randomly sampled from the corresponding domain item set together with the positive item as the final candidates. We report the recommendation performance on two widely used metrics, i.e., Hit Ratio, and NDCG [43] at cutoff k (**HR@k and NDCG@k**).

5.2.3 Implementation Details. In our framework, we first adopted the open-sourced Baichuan2-7B-Chat[3] model as the LLM to generate cross-domain open recommendations. Then, the text embedding model used in our work was chosen as the jina-embeddings-v2-base-en[4] model. In terms of the hyper-parameters, the dimensionality d of all latent representations was set to 64. The batch size was set to 128. The learning rate of the recommenders was set to 0.001. In the contrastive pre-training stage, the co-efficient λ_c of the contrastive loss was set to 0.1. The self-attention network was used as the sequential encoder in our work, and two single-head blocks were adopted as the original paper [36] suggested. Besides, we also performed a random search of a few hyper-parameters on the validation set, i.e., the number of recalled items K in $\{5, 10, 15, 20\}$, the meta update learning rate η_2 in $\{10^{-3}, 10^{-4}, 10^{-5}\}$, and the meta update interval Δu in $\{1, 10, 30, 50\}$. As for the baselines, we used the recommended hyper-parameters for each method. We also selected binary cross-entropy as the loss function, consistent with our approach for a fair comparison.

5.3 Overall Performance (RQ1)

We conducted experiments in Movie-Book and Food-Kitchen CDSR scenarios to evaluate the effectiveness of our LLMCDSR framework. The overall performance of LLMCDSR, compared to baseline models, is displayed in Tables 3 and 4 for the Movie-Book and Food-Kitchen scenarios, respectively, where the mean results of the metrics are obtained through three random runs using different initial parameters. The best and second-best results are highlighted in bold and underlined, respectively. LLMCDSR demonstrates superior performance in all metrics for the Movie-Book scenario and nearly all metrics for the Food-Kitchen scenario, indicating its effectiveness in utilizing non-overlapped information to improve CDSR. Specifically, in the Movie-Book scenario, LLMCDSR shows an average improvement of 20.78% and 14.88% in NDCG@k and HR@k for Movie recommendations, and 18.57% and 20.55% for Book recommendations. In the Food-Kitchen scenario, LLMCDSR outperforms the baselines in 10 out of 12 metrics, with average improvements of 2.31% and 4.36% in NDCG@5/10 and HR@k for Food recommendations, and 3.27% and 1.70% in NDCG@k and HR@3/5 for Kitchen recommendations. Additionally, our analysis yields several observations.

Firstly, it is noteworthy that CDSR methods like TPUF, π-Net, C2DSR, and MGCL exhibit surprisingly unsatisfactory performance, even falling short of single-domain recommendation methods. This primarily stems from their heavy reliance on overlapped users. In our more general

[3] https://huggingface.co/baichuan-inc/Baichuan2-7B-Chat
[4] https://huggingface.co/jinaai/jina-embeddings-v2-base-en

Table 3. The Overall Recommendation Performance on the Movie-Book Scenario

Methods	Movie						Book					
	NDCG			HR			NDCG			HR		
	@3	@5	@10	@3	@5	@10	@3	@5	@10	@3	@5	@10
TPUF	0.0189	0.0284	0.0346	0.0257	0.0482	0.0675	0.0148	0.0211	0.0287	0.0190	0.0343	0.0590
π-Net	0.0223	0.0281	0.0385	0.0302	0.0444	0.0772	0.0266	0.0332	0.0443	0.0377	0.0537	0.0887
C2DSR	0.0237	0.0282	0.0349	0.0322	0.0431	0.0637	0.0070	0.0103	0.0143	0.0107	0.0187	0.0312
MGCL	0.0278	0.0334	0.0471	0.0418	0.0547	0.0965	0.0215	0.0271	0.0362	0.0290	0.0427	0.0709
FPMC	0.0175	0.0219	0.0305	0.0232	0.0341	0.0617	0.0425	0.0500	0.0610	0.0545	0.0728	0.1078
Caser	0.0205	0.0249	0.0310	0.0270	0.0379	0.0572	0.0328	0.0403	0.0531	0.0438	0.0621	0.1021
FEARec	0.0148	0.0210	0.0297	0.0199	0.0354	0.0624	0.0442	0.0520	0.0631	0.0552	0.0743	0.1086
GRU4Rec	0.0188	0.0255	0.0341	0.0270	0.0431	0.0701	0.0445	0.0540	0.0675	0.0602	0.0838	0.1257
SRGNN	0.0209	0.0279	0.0369	0.0277	0.0450	0.0727	0.0248	0.0320	0.0417	0.0324	0.0499	0.0804
SASRec	0.0185	0.0230	0.0322	0.0257	0.0367	0.0650	0.0637	0.0797	0.1018	0.0823	0.1211	0.1893
LLMCDSR	**0.0339**	**0.0415**	**0.0547**	**0.0474**	**0.0659**	**0.1069**	**0.0766**	**0.0963**	**0.1167**	**0.1024**	**0.1505**	**0.2138**

Table 4. The Overall Recommendation Performance on the Food-Kitchen Scenario

Methods	Food						Kitchen					
	NDCG			HR			NDCG			HR		
	@3	@5	@10	@3	@5	@10	@3	@5	@10	@3	@5	@10
TPUF	0.0071	0.0098	0.0174	0.0100	0.0167	0.0400	0.0065	0.0090	0.0124	0.0078	0.0141	0.0251
π-Net	0.0182	0.0235	0.0321	0.0233	0.0367	0.0627	0.0167	0.0239	0.0357	0.0235	0.0411	0.0782
C2DSR	0.0438	0.0525	0.0688	0.0580	0.0793	0.1307	0.0131	0.0164	0.0227	0.0182	0.0264	0.0462
MGCL	0.0204	0.0284	0.0412	0.0253	0.0447	0.0853	0.0236	0.0290	0.0412	0.0301	0.0433	0.0819
FPMC	0.0851	0.1006	0.1258	0.1127	0.1507	0.2287	0.0582	0.0701	0.0857	0.0750	0.1043	0.1526
Caser	0.0267	0.0378	0.0580	0.0433	0.0700	0.1333	0.0517	0.0639	0.0847	0.0722	0.1020	0.1664
FEARec	0.0288	0.0343	0.0522	0.0367	0.0500	0.1067	0.0436	0.0534	0.0675	0.0571	0.0810	0.1250
GRU4Rec	0.0720	0.0896	0.1180	0.1006	0.1433	0.2320	0.0807	0.1013	0.1295	0.1086	0.1589	0.2468
SRGNN	0.0684	0.0862	0.1080	0.0956	0.1392	0.2072	0.0661	0.0811	0.0997	0.0861	0.1224	0.1807
SASRec	**0.1348**	0.1672	0.2125	0.1822	0.2600	0.4000	0.1698	0.2045	0.2488	0.2276	0.3124	**0.4495**
LLMCDSR	0.1322	**0.1684**	**0.2208**	**0.1833**	**0.2708**	**0.4333**	**0.1764**	**0.2138**	**0.2522**	**0.2296**	**0.3203**	0.4388

scenario, which includes both overlapped and non-overlapped users, the latter group constitutes a larger proportion of the overall user base. In these CDSR methods, knowledge transfer across domains relies on correlating interaction patterns from different domains based on overlapped user behaviors. However, this mechanism of transfer can be underdeveloped due to limited cross-domain behaviors. Furthermore, these methods fail to effectively incorporate single-domain interactions when there is no specific information available about the source or target domain. Remarkably, the underperformance of these methods can also be explained from the perspective of the data sparsity problem commonly found in recommender systems. If single-domain interactions are not effectively utilized, this issue is exacerbated within the limited cross-domain interactions. Specifically, as shown in Table 2, the original data densities for Movie-Book and Food-Kitchen are only 0.0099% and 0.0226%, respectively. However, these densities drop dramatically to 0.0007% and 0.0046% when only overlapped users are effectively considered, marking declines of 92.93% and 79.65%. Recommendation models typically face difficulties in learning valid collaborative knowledge in such sparse environments.

Secondly, single-domain SR methods show promising performance in CDSR tasks involving both overlapped and non-overlapped users. These methods treat items from multiple domains as if they belong to a single domain, modeling sequences without distinguishing between domains. Consequently, they utilize the abundant single-domain sequences more effectively compared to CDSR methods that handle domains separately and lack knowledge transfer for these single-domain sequences. However, a limitation of these methods is the insufficient utilization of single-domain interactions, as they do not facilitate explicit knowledge transfer across domains. Additionally, we observe that SASRec outperforms other single-domain models like GRU4Rec and SRGNN. This could be attributed to the mixed sequences from multiple domains not strictly adhering to ordinal properties, which may lead to biased results in models focused on capturing transition patterns. In contrast, SASRec, employing a self-attention architecture, is capable of capturing higher-order and wider-range dependencies through its full attention mechanism. This effectiveness of self-attention in handling diverse dependencies is also why we chose a self-attention-based sequential encoder in our framework.

Thirdly, our framework demonstrates more improvements over baselines in the Movie-Book dataset compared to the Food-Kitchen dataset. This discrepancy likely stems from the varying levels of difficulty in inferring cross-domain behaviors in these two scenarios. The simpler it is to infer these behaviors, the more reliable the LLM-generated content becomes, thereby more effectively enhancing CDSR. Specifically, in the Movie-Book scenario, correlations between the two domains are relatively easier to discern. For instance, a user's specific preferences in movies often align with their book choices. A user interested in science fiction movies is likely to enjoy similar genres in books. Moreover, many movies and books are part of the same series (*e.g.,* "Harry Potter" or "The Great Gatsby"), and users often follow these series across both domains, simplifying cross-domain inferences. In contrast, in the Food-Kitchen scenario, food preferences do not always directly correlate with kitchenware preferences, making it more difficult to draw reliable inferences between these domains. Nevertheless, owing to the exceptional capabilities of LLMs in grasping cross-domain knowledge, the performance in the Food-Kitchen scenario is still notably improved when utilizing our LLMCDSR framework.

5.4 Ablation Study (RQ2)

To evaluate the effectiveness of different components within the LLMCDSR framework, we conduct a comparative analysis with several variants of LLMCDSR:

- *LLMCDSR-text* retrieves the pseudo-cross-domain items based solely on the cosine similarity between the textual embeddings of the items from the dataset and the LLM generations.
- *LLMCDSR-random* randomly samples cross-domain items from the dataset to form the pseudo-cross-domain sequences.
- *LLMCDSR-meta* only excludes the RMRN, such that the ability to adaptively recall cross-domain items is deprived and the selection of pseudo cross-domain items remains static and unadjusted for relevance.
- *LLMCDSR-proj* removes the use of pre-trained textual projection matrices, such that no collaborative information is infused into the textual representations.
- *LLMCDSR-ind* removes the indicator function in the training loss of the unified recommender such that the pseudo items are included as target items during the training process.

As can be seen in Figure 4, LLMCDSR outshines most of its variants, highlighting the effectiveness of our design elements in addressing the CDSR. In particular, it is observed that directly incorporating the candidate-free generation results based on textual similarity from LLMs (LLMCDSR-text) generally yields suboptimal performance. This is attributed to the raw recall strategy's lack of

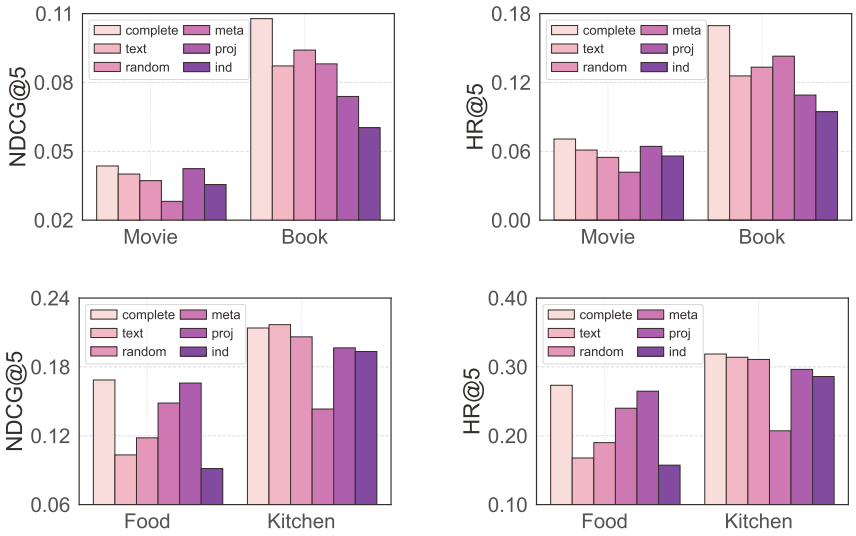
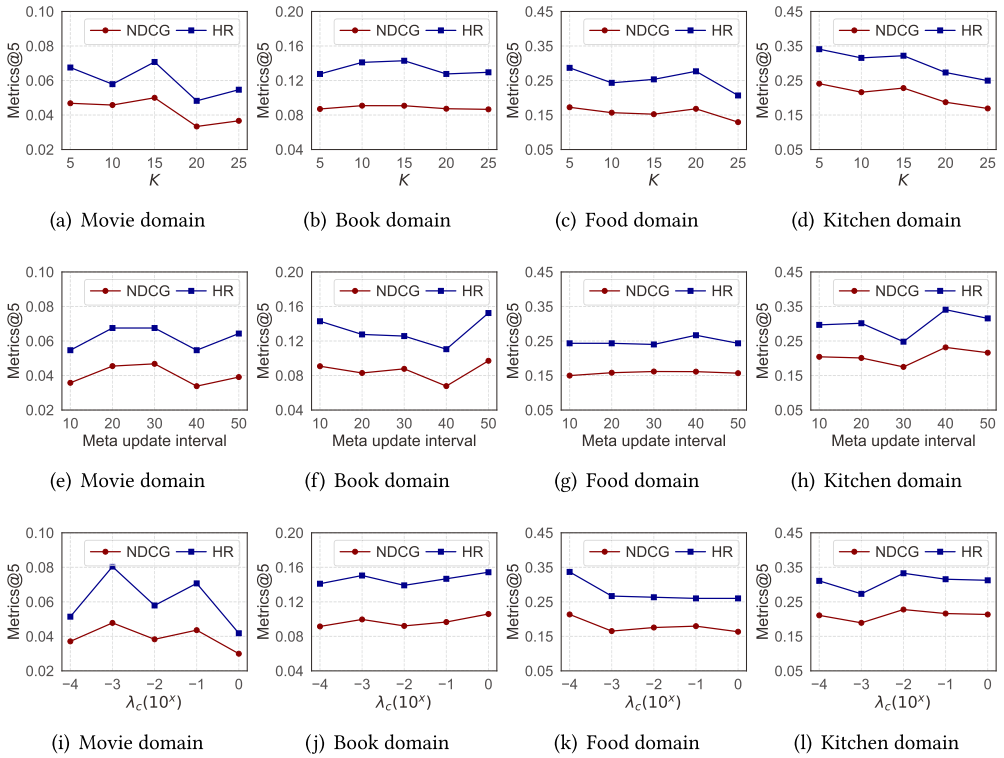


Fig. 4. Contributions of different components in LLMCDSR.

collaborative signal. Additionally, the content generated by LLMs may include less relevant and important elements, potentially detrimental to CDSR performance. Interestingly, in specific instances, such as within the Food domain, directly using retrieved items based on textual similarity delivers promising results. This could be due to relatively lower noise levels in the textual generations within this domain. Furthermore, our analysis reveals that in certain cases, training with items recalled by textual similarity is less effective than using randomly sampled cross-domain items (LLMCDSR-random). This finding suggests that the original generation results from LLMs are not invariably beneficial. At times, they may even be less effective than random selection, emphasizing the need for careful curation and adjustment of generated content to optimize recommendations. We also observe that augmenting single-domain sequences with random items outperforms many baselines. This gap occurs because the baselines typically suffer from inadequate cross-domain indicators, resulting in both sparsity issues and poor cross-domain knowledge alignment. In contrast, randomly adding cross-domain items can help mitigate this problem. However, this approach does not match the performance of the complete LLMCDSR since the randomly injected items may be of low quality and fail to capture the true cross-domain distributions. Besides, incorporating pseudo items as target items to predict the loss function (LLMCDSR-ind) leads to a performance decline. This happens because treating these augmented items as positive items can introduce noise or bias into the training process, which confirms the essential role of the proposed indicator function.

The consistent performance drop when RMRN is removed, as in LLMCDSR-meta, highlights the importance of relevance weighting in the generation samples. Additionally, the performance of LLMCDSR-proj demonstrates the positive effect of collaborative information, as learned through collaborative-textual contrastive pre-training, affirming its value in our framework. Notably, this collaborative information more substantially benefits the Book and Kitchen domains compared to the Movie and Food domains. This difference is likely due to the greater volume of interactions in the former, as detailed in Table 2. The abundance of data in these interaction-rich domains allows for more effective capture and utilization of collaborative signals.



(a) Movie domain (b) Book domain (c) Food domain (d) Kitchen domain

(e) Movie domain (f) Book domain (g) Food domain (h) Kitchen domain

(i) Movie domain (j) Book domain (k) Food domain (l) Kitchen domain

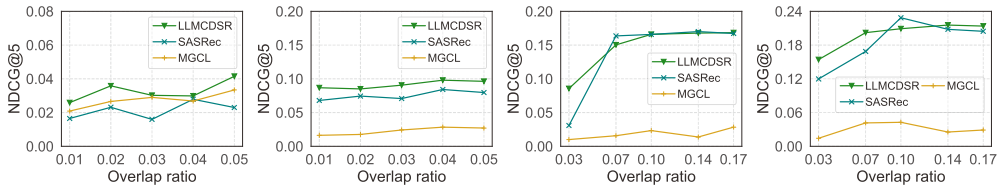
Fig. 5. Parameter sensitivity analysis on four domains of two datasets.

5.5 Parameter Sensitivity (RQ3)

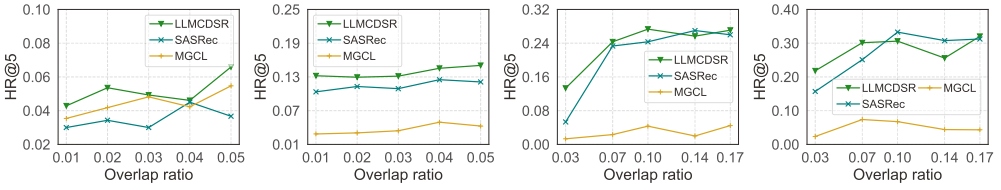
Figure 5 presents the parameter sensitivity on four recommendation domains in the two datasets. The analysis of the two studied parameters is conducted as follows.

Number of Recalled Items K. We analyze how the number of recalled cross-domain items in Equation (14) affects the performance by varying K from 5 to 25. As shown in Figure 5(a) to (d), for the Movie-Book dataset, we observed an increase in performance as K rose to 15, followed by a decline. Conversely, in the Food-Kitchen dataset, performance consistently decreased with an increase in K. The parameter K dictates the quantity of pseudo-cross-domain items within the pseudo-cross-domain sequences. While these pseudo items can effectively facilitate knowledge transfer across domains, they are not universally beneficial to the CDSR task. This is due to their estimated nature, which may not precisely mirror true cross-domain preferences. Consequently, it is crucial to strike a balance between the pseudo and real items in the sequences. The Movie-Book dataset, characterized by longer average interaction sequence lengths, appears to accommodate a greater number of pseudo items compared to the Food-Kitchen domain.

Meta Update Interval Δu. To further investigate the impact of the frequency for updating the meta parameter (*i.e.,* $\boldsymbol{w}$), we vary Δu from 10 to 50 on all domains. As can be seen in Figure 5(e)–(h), the performance peaks when Δu is relatively large. This trend suggests that less frequent updates of the meta parameter allow the training process to concentrate more on the primary recommendation task. Consequently, overly frequent updates of the meta parameter can potentially disrupt the main objective of the model, adversely affecting recommendation performance. Additionally, the bi-level optimization process, which entails model duplication and computation of second-order



(a) Movie domain, NDCG@5 (b) Book domain, NDCG@5 (c) Food domain, NDCG@5 (d) Kitchen domain, NDCG@5



(e) Movie domain, HR@5 (f) Book domain, HR@5 (g) Food domain, HR@5 (h) Kitchen domain, HR@5

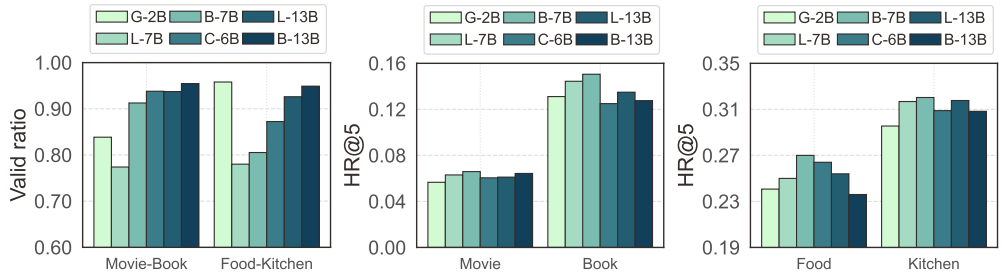
Fig. 6. The influence of overlap ratio on the CDSR performance.

gradients, becomes less efficient with too frequent meta updates. Therefore, finding an optimal Δu balance is crucial for both enhancing the model's recommendation accuracy and ensuring training efficiency.

Strength of the Contrastive Alignment λ_c. We also study how the strength of the contrastive alignment in the Collaborative-Textual Contrastive Pre-Training affects recommendation performance by varying λ_c from 10^{-4} to 1. As shown in Figure 5(i)–(l), the performance improves with a weaker contrastive alignment on the Movie and Food domains, whereas the opposite is true for the Book and Kitchen domains. This difference arises because the Movie and Food domains have less collaborative information available, due to fewer recorded interactions (see Table 2). Consequently, these domains can achieve good performance even with a weaker alignment, as there is less need to align collaborative knowledge.

5.6 Influence of Overlapped Users (RQ4)

As overlapped users play a crucial role in CDSR, we examined the impact of varying overlap ratios on the performance of LLMCDSR and other competitive single- and cross-domain baselines (e.g., MGCL and SASRec), adjusting the ratio from 0.01 to 0.05 for the Movie-Book dataset, and 0.03 to 0.17 for the Food-Kitchen dataset. The upper limits of these ranges correspond to the initial overlap ratios of their respective datasets. The findings, depicted in Figure 6, first reveal that MGCL performs poorly across all levels of user overlap in all domains except for Movies, underscoring its limitations in CDSR when overlap is minimal and non-overlapped users are prevalent. The possible reason for MGCL's acceptable performance in the Movie domain is that the patterns in this domain are effectively captured by the overlapping data, facilitating MGCL's smoother utilization of this information. Notably, LLMCDSR demonstrates resilience against the reduction of cross-domain information. This robustness is attributed to the use of estimated cross-domain interactions by LLMs, derived from original single-domain interactions, to enhance CDSR. Specifically, in the Movie-Book scenario, LLMCDSR consistently outperforms the other baselines at all user overlap ratios. In the Food-Kitchen scenario, while LLMCDSR and SASRec show similar performance at higher overlap ratios, LLMCDSR significantly excels at lower levels (e.g., 0.03), displaying greater resistance to decreases in the overlap ratio. This confirms the minimal reliance of our framework on overlapped users and its superior capability in leveraging available non-overlapped data for CDSR. Additionally,



(a) The ratio of valid generation samples. (b) Movie-Book (c) Food-Kitchen

Fig. 7. The study of different LLM backbones including Gemma2-2B(G-2B), Llama2-7/13B(L-7/13B), Baichuan2-7/13B(B-7/13B), and ChatGLM3-6B(C-6B).

we observe occasional performance dips at certain points of increased overlap ratio. This could be attributed to the random sampling of added overlapped users introducing redundant noise, adversely affecting performance. However, from a broader perspective, an increase in overlapped information generally contributes positively to CDSR performance.

5.7 LLM Backbone Study (RQ5)

To probe the effectiveness of different LLMs with different model sizes for generating cross-domain interactions in the CDSR, we select three prevalent LLMs, i.e., gemma-2-2b-it[5], Llama-2-7b-chat[6], Llama-2-13b-chat[7], Baichuan2-7B-Chat, Baichuan2-13B-Chat[8] and chatglm3-6b[9] as the cross-domain interaction generator in our framework. We only retain valid generation results for each single-domain sequence input, and the valid ratio for each LLM on two datasets is illustrated in Figure 7(a). The performances using different LLM backbones in the Movie-Book scenario and the Food-Kitchen scenario are shown in Figure 7(b) and (c), respectively. The results indicate that larger models generally produce more valid generations, reflecting their superior ability to follow instructions. Interestingly, Gemma2-2B also generates a substantial number of valid outputs, suggesting it can effectively understand instructions even with fewer parameters. However, in terms of recommendation performance, we observe that Gemma2-2B does not achieve competitive results, demonstrating its limitations in performing the cross-domain interaction generation task. Additionally, larger models like Baichuan2-13B, which produce more valid generations, do not necessarily achieve better performance. This suggests that their generated interactions may contain potential undesired noise irrelevant to our task. These findings highlight a tradeoff between the quality and quantity of generated content in our task. As a result, we finally selected Baichuan2 as the LLM backbone for our framework due to its balanced combination of strong performance and efficient inference. We also notice a higher occurrence of valid generations in the Movie-Book scenario compared to the Food-Kitchen scenario. This is likely because more closely related domains make it easier for LLMs to generate meaningful cross-domain interactions.

[5]https://huggingface.co/google/gemma-2-2b-it
[6]https://huggingface.co/meta-llama/Llama-2-7b-chat-hf
[7]https://huggingface.co/meta-llama/Llama-2-13b-chat-hf
[8]https://huggingface.co/baichuan-inc/Baichuan2-13B-Chat
[9]https://huggingface.co/THUDM/chatglm3-6b

Table 5. The Efficiency Comparison in Terms of Training (s/epoch) and Inference ($10^{-3}s$/Instance) Time

Methods	Movie-Book		Food-Kitchen	
	Training	Inference	Training	Inference
TPUF	1.15/8.09	2.98/3.71	0.77/3.83	3.90/4.76
π-Net	315.65	10.32	80.39	36.15
C2DSR	3.33	3.18	1.64	3.03
MGCL	225.74/256.09	7.89	86.08/93.12	6.90
FPMC	70.87	0.39	12.92	0.40
Caser	1562.32	7.48	416.73	8.38
FEARec	343.08	12.50	90.50	4.21
GRU4Rec	45.19	0.41	12.64	0.41
SRGNN	435.33	1.14	120.41	2.08
SASRec	11.14	2.34	3.26	2.25
Pre-training	2.37/21.32	-	1.81/6.16	-
RMRN	42.27	3.54	4.88	1.94

The time for different domains is reported separately if they require distinct training or inference processes.

5.8 Efficiency Evaluation (RQ6)

We also conduct experiments to evaluate the efficiency of our framework. Specifically, we compare the Collaborative-Textual Contrastive Pre-training and the RMRN module in our LLMCDSR against the baselines in terms of the average time to train one epoch and the average inference time per instance. Note that we do not report the time required for cross-domain interaction generation by LLMs since it can be performed offline beforehand. For a fair comparison, the batch size is set to 256 for all models in these experiments. As shown in Table 5, our methods outperform most baselines in training time, demonstrating the efficiency of our proposed framework. While a few baselines are more efficient, they do not achieve the same level of recommendation performance, making the modest efficiency tradeoff acceptable. Furthermore, since LLMCDSR incorporates only a basic recommender for inference, it achieves low inference latency, confirming its practicality for real-world applications.

5.9 Effectiveness of RMRN (RQ7)

To validate the effectiveness of the RMRN process in our framework, we conduct both quantitative and qualitative analyses.

5.9.1 Quantitative Analysis. To assess the quality of the pseudo items recalled by RMRN, we first use LLMs to generate cross-domain interactions for overlapped users who have actual cross-domain interactions in the test set. We then compare the Euclidean distance between the anchor representation, generated with weighted relevance by RMRN (Equation (13)), and the representation of the user's real interactions against the distance obtained using an aggregated representation with vanilla identical relevance. As shown in Table 6, the anchor representation with weighted relevance is closer to the representation of users' actual interactions. This improved alignment indicates that the representations generated by our framework facilitate the recall of pseudo-items that more

Table 6. Euclidean Distance between Anchor and Real Interaction Representations Using Two Different Weighting Strategies for Overlapped Users in the Test Set

Weighting strategy	Movie-Book		Food-Kitchen	
	Movie	Book	Food	Kitchen
Identical relevance	2.44	10.14	2.60	1.31
Weighted relevance (RMRN)	**1.36**	**9.23**	**1.49**	**0.39**

The best results are in bold.

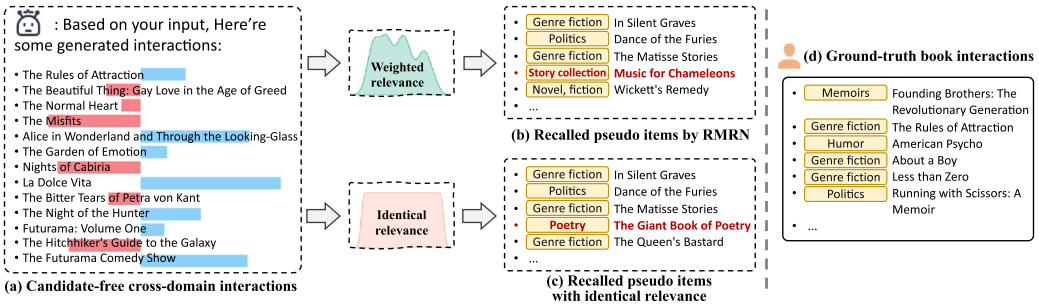


Fig. 8. The illustration of the case study.

accurately reflect real cross-domain distributions, thus enhancing cross-domain knowledge transfer and promoting CDSR.

5.9.2 Qualitative Analysis. We select a random overlapped user from the test set of the Movie-Book dataset, with the ground-truth Book interactions shown in Figure 8(d). Then, using LLMCDSR, we generate the cross-domain interactions (Figure 8(a)), with the relevance of each generation indicated by blue and red bars. The further to the right, the more positive the relevance. Afterwards, we recall pseudo Book items for this user based on the Movie interactions, listed in Figure 8(b). For comparison, we also performed a recall using a vanilla approach with equal relevance for each sample, with results shown in Figure 8(c).

LLMs' raw generations show a diverse range of book titles, some of which differ significantly from items in the dataset. The recall operation, informed by weighted relevance, enables retrieving pseudo items that are more closely aligned with the dataset. Secondly, Some generated titles, like "The Beautiful Thing" and "Nights of Cabiria," are better known as films rather than books, likely reflecting the hallucination issue inherent in the LLMs. Our relevance weighting layer effectively downgrades the weight of such samples, benefiting the recall and recommendation process. Thirdly, the pseudo items recalled by our framework closely correspond with the user's ground-truth preferences. This includes a similar mix of genre fiction and politics books as found in the user's actual interactions. This alignment suggests LLMs' capability to infer latent cross-domain interactions, thereby effectively aiding the CDSR. Finally, a comparison between pseudo items recalled with identical relevance and those with RMRN's weighted relevance highlights a difference where RMRN tends to favor story collections over poetry. This preference aligns more closely with the user's cross-domain interactions, suggesting the user is likely a "prose reader," as inferred from the ground-truth interactions.

6 Conclusion and Future Work

In this article, we explored the CDSR problem in scenarios involving both overlapping and non-overlapping users. We proposed LLMCDSR to tackle the problem by leveraging LLMs to estimate the pseudo-cross-domain items in single-domain interactions, thereby enhancing knowledge alignment and transfer across domains. Along this line, we first prompted LLMs to generate text-like cross-domain interactions without a candidate selection procedure. Then, we devised a Collaborative-Textual Contrastive Pre-Training strategy to learn how to effectively infuse the collaborative information into the textual features. Furthermore, a novel RMRN was proposed to automatically identify and retrieve high-relevance pseudo-cross-domain items w.r.t. the downstream task, with parameter optimization achieved through an advanced bi-level meta-learning. Comprehensive experiments across four recommendation domains in two real-world public datasets affirm the effectiveness of LLMCDSR.

In future work, we aim to explore fine-tuning techniques to improve the quality of LLM-generated data from the source, which could further enhance CDSR performance. Additionally, we plan to investigate the integration of multiple source domains for the target domain, as this more general setting could lead to further improvements in CDSR performance.

References

[1] Taha Aksu, Min-Yen Kan, and Nancy F. Chen. 2023. Prompter: Zero-shot adaptive prefixes for dialogue state tracking domain adaptation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers. Retrieved from https://doi.org/10.18653/v1/2023.acl-long.252

[2] Keqin Bao, Jizhi Zhang, Wenjie Wang, Yang Zhang, Zhengyi Yang, Yancheng Luo, Chong Chen, Fuli Feng, and Qi Tian. 2023. A bi-step grounding paradigm for large language models in recommendation systems. arXiv:2308.08434.

[3] 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 *RecSys '23: Proceedings of the 17th ACM Conference on Recommender Systems*, 1007–1014. Retrieved from https://doi.org/10.1145/3604915.3608857

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 33, 1877–1901.

[5] Jiangxia Cao, Xin Cong, Jiawei Sheng, Tingwen Liu, and Bin Wang. 2022. Contrastive cross-domain sequential recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 138–147.

[6] Huiyuan Chen, Yusan Lin, Menghai Pan, Lan Wang, Chin-Chia Michael Yeh, Xiaoting Li, Yan Zheng, Fei Wang, and Hao Yang. 2022. Denoising self-attentive sequential recommendation. In *Proceedings of the 16th ACM Conference on Recommender Systems*, 92–101.

[7] Liyi Chen, Panrong Tong, Zhongming Jin, Ying Sun, Jieping Ye, and Hui Xiong. 2024. Plan-on-graph: Self-correcting adaptive planning of large language model on knowledge graphs. arXiv:2410.23875. Retrieved from https://doi.org/10.48550/arXiv.2410.23875

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the International Conference on Machine Learning*. PMLR, 1597–1607.

[9] Xinlei Chen and Kaiming He. 2021. Exploring simple siamese representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15750–15758.

[10] Mingyue Cheng, Fajie Yuan, Qi Liu, Xin, and Enhong Chen. 2021. Learning transferable user representations with sequential behaviors via contrastive pre-training. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '21)*. IEEE, 51–60.

[11] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACM, 1724–1734. DOI: https://doi.org/10.3115/v1/D14-1179

[12] Sunhao Dai, Ninglu Shao, Haiyuan Zhao, Weijie Yu, Zihua Si, Chen Xu, Zhongxiang Sun, Xiao Zhang, and Jun Xu. 2023. Uncovering ChatGPT's capabilities in recommender systems. In *RecSys '23: Proceedings of the 17th ACM Conference on Recommender Systems*, 1126–1132. DOI: https://doi.org/10.1145/3604915.3610646

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1: Long and Short Papers, ACM, 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423

[14] Yujia Ding, Huan Li, Ke Chen, and Lidan Shou. 2023. TPUF: Enhancing cross-domain sequential recommendation via transferring pre-trained user features. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, 410–419.

[15] Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, and Zhifang Sui. 2024. A survey for in-context learning. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. ACM, 1107–1128. DOI: https://doi.org/10.18653/v1/2024.emnlp-main.64

[16] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. 2017. Sequential user-based recurrent neural network recommendations. In *Proceedings of the 11th ACM Conference on Recommender Systems*, 152–160.

[17] Xinyu Du, Huanhuan Yuan, Pengpeng Zhao, Jianfeng Qu, Fuzhen Zhuang, Guanfeng Liu, Yanchi Liu, and Victor S Sheng. 2023. Frequency enhanced hybrid attention network for sequential recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 78–88.

[18] Ziwei Fan, Zhiwei Liu, Yu Wang, Alice Wang, Zahra Nazari, Lei Zheng, Hao Peng, and Philip S Yu. 2022. Sequential recommendation via stochastic self-attention. In *Proceedings of the ACM Web Conference 2022*, 2036–2047.

[19] Chuyu Fang, Chuan Qin, Qi Zhang, Kaichun Yao, Jingshuai Zhang, Hengshu Zhu, Fuzhen Zhuang, and Hui Xiong. 2023. Recruitpro: A pretrained language model with skill-aware prompt learning for intelligent recruitment. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3991–4002.

[20] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the International Conference on Machine Learning*. PMLR, 1126–1135.

[21] Jiahui Gao, Renjie Pi, Lin Yong, Hang Xu, Jiacheng Ye, Zhiyong Wu, Weizhong Zhang, Xiaodan Liang, Zhenguo Li, and Lingpeng Kong. 2023. Self-guided noise-free data generation for efficient zero-shot learning. In *International Conference on Learning Representations (ICLR 2023)*.

[22] Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2022. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM Conference on Recommender Systems*. 299–315.

[23] Lei Guo, Li Tang, Tong Chen, Lei Zhu, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2021. DA-GCN: A domain-aware attentive graph convolution network for shared-account cross-domain sequential recommendation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence Main Track*, 2483–2489. DOI: https://doi.org/10.24963/ijcai.2021/342

[24] Yongjing Hao, Tingting Zhang, Pengpeng Zhao, Yanchi Liu, Victor S Sheng, Jiajie Xu, Guanfeng Liu, and Xiaofang Zhou. 2023. Feature-level deeper self-attention network with contrastive learning for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 10 (2023), 10112–10124.

[25] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *Proceedings of the IEEE 16th International Conference on Data Mining (ICDM '16)*. IEEE, 191–200.

[26] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, Yongdong Zhang, and Meng Wang. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 639–648.

[27] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, 173–182.

[28] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. 2015. Session-based recommendations with recurrent neural networks. arXiv:1511.06939. Retrieved from https://doi.org/10.48550/arXiv.1511.06939

[29] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[30] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. Conet: Collaborative cross networks for cross-domain recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 667–676.

[31] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. 2013. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd International Conference on World Wide Web*, 595–606.

[32] Chao Huang, Xian Wu, Xuchao Zhang, Chuxu Zhang, Jiashu Zhao, Dawei Yin, and Nitesh V Chawla. 2019. Online purchase prediction via multi-scale modeling of behavior dynamics. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2613–2622.

[33] Jianchao Ji, Zelong Li, Shuyuan Xu, Wenyue Hua, Yingqiang Ge, Juntao Tan, and Yongfeng Zhang. 2024. Genrec: Large language model for generative recommendation. In *Advances in Information Retrieval ECIR 2024*. Lecture Notes in Computer Science, Vol. 14610, Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-56063-7_42

[34] Meng Jiang, Peng Cui, Nicholas Jing Yuan, Xing Xie, and Shiqiang Yang. 2016. Little is much: Bridging cross-platform behaviors through overlapped crowds. *In Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[35] Wei Jin, Haitao Mao, Zheng Li, Haoming Jiang, Chen Luo, Hongzhi Wen, Haoyu Han, Hanqing Lu, Zhengyang Wang, Ruirui Li, et al. 2023. Amazon-m2: A multilingual multi-locale shopping session dataset for recommendation and text generation. arXiv:2307.09688.

[36] Wang-Cheng Kang and Julian McAuley. 2018. Self-attentive sequential recommendation. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '18)*. IEEE, 197–206.

[37] Wang-Cheng Kang, Jianmo Ni, Nikhil Mehta, Maheswaran Sathiamoorthy, Lichan Hong, Ed Chi, and Derek Zhiyuan Cheng. 2023. Do LLMs Understand user preferences? Evaluating LLMs on user rating prediction. arXiv:2305.06474. Retrieved from https://doi.org/10.48550/arXiv.2305.06474

[38] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907. Retrieved from https://doi.org/10.48550/arXiv.1609.02907

[39] Pan Li and Alexander Tuzhilin. 2021. Dual metric learning for effective and efficient cross-domain recommendations. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2021), 321–334.

[40] Shuangli Li, Jingbo Zhou, Tong Xu, Dejing Dou, and Hui Xiong. 2022. Geomgcl: Geometric graph contrastive learning for molecular property prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. 4541–4549.

[41] Feng Liang, Weike Pan, and Zhong Ming. 2021. Fedrec++: Lossless federated recommendation with explicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4224–4231.

[42] Jiayi Liao, Sihang Li, Zhengyi Yang, Jiancan Wu, Yancheng Yuan, Xiang Wang, and Xiangnan He. 2024. Llara: Large language-recommendation assistant. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1785–1795.

[43] Guanyu Lin, Chen Gao, Yu Zheng, Jianxin Chang, Yanan Niu, Yang Song, Kun Gai, Zhiheng Li, Depeng Jin, Yong Li, et al. 2024. Mixed Attention Network for Cross-domain Sequential Recommendation. In *WSDM '24: Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 405–413. DOI: https://doi.org/10.1145/3616855.3635801

[44] Jianghao Lin, Xinyi Dai, Rong Shan, Bo Chen, Ruiming Tang, Yong Yu, and Weinan Zhang. 2025. Large language models make sample-efficient recommender systems. *Frontiers of Computer Science* 19, 4 (2025), 194328.

[45] Xinyu Lin, Wenjie Wang, Yongqi Li, Fuli Feng, See-Kiong Ng, and Tat-Seng Chua. 2023. A multi-facet paradigm to bridge large language model and recommendation. arXiv:2310.06491. Retrieved from https://doi.org/10.48550/arXiv.2310.06491

[46] Junling Liu, Chao Liu, Renjie Lv, Kang Zhou, and Yan Zhang. 2023. Is ChatGPT a good recommender? A preliminary study. arXiv:2304.10149. Retrieved from https://doi.org/10.48550/arXiv.2304.10149

[47] Jian Liu, Pengpeng Zhao, Fuzhen Zhuang, Yanchi Liu, Victor S Sheng, Jiajie Xu, Xiaofang Zhou, and Hui Xiong. 2020. Exploiting aesthetic preference in deep cross networks for cross-domain recommendation. In *Proceedings of The Web Conference 2020*, 2768–2774.

[48] Kang Liu, Feng Xue, Dan Guo, Le Wu, Shujie Li, and Richang Hong. 2023. Megcf: Multimodal entity graph collaborative filtering for personalized recommendation. *ACM Transactions on Information Systems* 41, 2 (2023), 1–27.

[49] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 885–894.

[50] Yan-Fu Liu, Cheng-Yu Hsu, and Shan-Hung Wu. 2015. Non-linear cross-domain collaborative filtering via hyper-structure transfer. In *Proceedings of the International Conference on Machine Learning*. PMLR, 1190–1198.

[51] Zhiwei Liu, Yongjun Chen, Jia Li, Philip S Yu, Julian McAuley, and Caiming Xiong. 2021. Contrastive self-supervised sequential recommendation with robust augmentation. arXiv:2108.06479. Retrieved from https://doi.org/10.48550/arXiv.2108.06479

[52] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.

[53] Chen Ma, Liheng Ma, Yingxue Zhang, Jianing Sun, Xue Liu, and Mark Coates. 2020. Memory augmented graph neural networks for sequential recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, 5045–5052.

[54] Huan Ma, Changqing Zhang, Yatao Bian, Lemao Liu, Zhirui Zhang, Peilin Zhao, Shu Zhang, Huazhu Fu, Qinghua Hu, and Bingzhe Wu. 2023. Fairness-guided few-shot prompting for large language models. arXiv:2303.13217. Retrieved from https://doi.org/10.48550/arXiv.2303.13217

[55] Muyang Ma, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Lifan Zhao, Peiyu Liu, Jun Ma, and Maarten de Rijke. 2022. Mixed information flow for cross-domain sequential recommendations. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16, 4 (2022), 1–32.

[56] Muyang Ma, Pengjie Ren, Yujie Lin, Zhumin Chen, Jun Ma, and Maarten de Rijke. 2019. π-net: A parallel information-sharing network for shared-account cross-domain sequential recommendations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 685–694.

[57] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-domain recommendation: An embedding and mapping approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI '17)*, Vol. 17, 2464–2470.

[58] Gustavo Penha and Claudia Hauff. 2020. What does Bert know about books, movies and music? Probing Bert for conversational recommendation. In *Proceedings of the 14th ACM Conference on Recommender Systems*, 388–397.

[59] Zhaopeng Qiu, Xian Wu, Jingyue Gao, and Wei Fan. 2021. U-BERT: Pre-training user representations for improved recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, 4320–4327.

[60] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the 11th ACM Conference on Recommender Systems*, 130–137.

[61] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*. PMLR, 8748–8763.

[62] Dimitrios Rafailidis and Fabio Crestani. 2017. A collaborative ranking model for cross-domain recommendations. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2263–2266.

[63] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research* 21, 1 (2020), 5485–5551.

[64] 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 on Web Conference 2024*, 3464–3475.

[65] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, 811–820.

[66] Scott Sanner, Krisztian Balog, Filip Radlinski, Ben Wedin, and Lucas Dixon. 2023. Large language models are competitive near cold-start recommenders for language-and item-based preferences. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 890–896.

[67] Guy Shani, David Heckerman, Ronen I Brafman, and Craig Boutilier. 2005. An MDP-based recommender system. *Journal of Machine Learning Research* 6, 9 (2005), 1265–1295.

[68] Jun Shu, Qi Xie, Lixuan Yi, Qian Zhao, Sanping Zhou, Zongben Xu, and Deyu Meng. 2019. Meta-weight-net: Learning an explicit mapping for sample weighting. In *Advances in Neural Information Processing Systems*, Vol. 32.

[69] Ajit P. Singh and Geoffrey J. Gordon. 2008. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 650–658.

[70] Wenchao Sun, Muyang Ma, Pengjie Ren, Yujie Lin, Zhumin Chen, Zhaochun Ren, Jun Ma, and Maarten De Rijke. 2021. Parallel split-join networks for shared account cross-domain sequential recommendations. *IEEE Transactions on Knowledge and Data Engineering* 35, 4 (2021), 4106–4123.

[71] Weiwei Sun, Lingyong Yan, Xinyu Ma, Pengjie Ren, Dawei Yin, and Zhaochun Ren. 2023. Is ChatGPT good at search? Investigating large language models as re-ranking agents. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. ACM, 14918–14937. DOI: https://doi.org/10.18653/v1/2023.emnlp-main.923

[72] Ying Sun, Yang Ji, Hengshu Zhu, Fuzhen Zhuang, Qing He, and Hui Xiong. 2024. Market-aware long-term job skill recommendation with explainable deep reinforcement learning. *ACM Transactions on Information Systems* 43, 2 (2024), 1–35.

[73] Ying Sun, Hengshu Zhu, Lu Wang, Le Zhang, and Hui Xiong. 2024. Large-scale online job search behaviors reveal labor market shifts amid COVID-19. *Nature Cities* 1, 2 (2024), 150–163.

[74] Jiaxi Tang and Ke Wang. 2018. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the eleventh ACM International Conference on Web Search and Data Mining*, 565–573.

[75] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv:2307.09288. Retrieved from https://doi.org/10.48550/arXiv.2307.09288

[76] Aaron Van Den Oord, and Oriol Vinyals. 2017. Neural discrete representation learning. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30.

[77] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30.

[78] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. arXiv:1710.10903. Retrieved from https://doi.org/10.48550/arXiv.1710.10903

[79] Chao Wang, Qi Liu, Runze Wu, Enhong Chen, Chuanren Liu, Xunpeng Huang, and Zhenya Huang. 2018. Confidence-aware matrix factorization for recommender systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[80] Chao Wang, Hengshu Zhu, Qiming Hao, Keli Xiao, and Hui Xiong. 2021. Variable interval time sequence modeling for career trajectory prediction: Deep collaborative perspective. In *Proceedings of the Web Conference 2021*, 612–623.

[81] Chao Wang, Hengshu Zhu, Peng Wang, Chen Zhu, Xi Zhang, Enhong Chen, and Hui Xiong. 2021. Personalized and explainable employee training course recommendations: A bayesian variational approach. *ACM Transactions on Information Systems (TOIS)* 40, 4 (2021), 1–32.

[82] Chao Wang, Hengshu Zhu, Chen Zhu, Chuan Qin, Enhong Chen, and Hui Xiong. 2023. Setrank: A setwise Bayesian approach for collaborative ranking in recommender system. *ACM Transactions on Information Systems* 42, 2 (2023), 1–32.

[83] Meirui Wang, Pengjie Ren, Lei Mei, Zhumin Chen, Jun Ma, and Maarten De Rijke. 2019. A collaborative session-based recommendation approach with parallel memory modules. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 345–354.

[84] Xinhua Wang, Houping Yue, Zizheng Wang, Liancheng Xu, and Jinyu Zhang. 2023. Unbiased and robust: External attention-enhanced graph contrastive learning for cross-domain sequential recommendation. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE. DOI: https://doi.org/10.1109/ICDMW60847.2023.00194

[85] Yau-Shian Wang, Ta-Chung Chi, Ruohong Zhang, and Yiming Yang. 2023. PESCO: Prompt-enhanced self contrastive learning for zero-shot text classification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, ACM, 14897–14911. DOI: https://doi.org/10.18653/v1/2023.acl-long.832

[86] Wei Wei, Xubin Ren, Jiabin Tang, Qinyong Wang, Lixin Su, Suqi Cheng, Junfeng Wang, Dawei Yin, and Chao Huang. 2024. Llmrec: Large language models with graph augmentation for recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 806–815.

[87] Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2022. UserBERT: Pre-training User Model with Contrastive Self-supervision. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2087–2092.

[88] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. 2019. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 346–353.

[89] Yunjia Xi, Weiwen Liu, Jianghao Lin, Jieming Zhu, Bo Chen, Ruiming Tang, Weinan Zhang, Rui Zhang, and Yong Yu. 2024. Towards open-world recommendation with knowledge augmentation from large language models. In *RecSys '24: Proceedings of the 18th ACM Conference on Recommender Systems*, 12–22. DOI: https://doi.org/10.1145/3640457.3688104

[90] Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. 2022. Contrastive learning for sequential recommendation. In *Proceedings of the IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 1259–1273.

[91] Xin Xin, Xiangyuan Liu, Hanbing Wang, Pengjie Ren, Zhumin Chen, Jiahuan Lei, Xinlei Shi, Hengliang Luo, Joemon M Jose, Maarten de Rijke, et al. 2023. Improving implicit feedback-based recommendation through multi-behavior alignment. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 932–941.

[92] Zitao Xu, Weike Pan, and Zhong Ming. 2023. A multi-view graph contrastive learning framework for cross-domain sequential recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 491–501.

[93] Bowen Yang, Cong Han, Yu Li, Lei Zuo, and Zhou Yu. 2022. Improving conversational recommendation systems' quality with context-aware item meta information. In *Findings of the Association for Computational Linguistics: NAACL 2022*. ACM, 38–48. DOI: https://doi.org/10.18653/v1/2022.findings-naacl.4

[94] Jia-Qi Yang, Chenglei Dai, Dan Ou, Dongshuai Li, Ju Huang, De-Chuan Zhan, Xiaoyi Zeng, and Yang Yang. 2025. COURIER: contrastive user intention reconstruction for large-scale visual recommendation. *Frontiers of Computer Science* 19, 7 (2025), 197602.

[95] Xiaoxin Ye, Yun Li, and Lina Yao. 2023. DREAM: Decoupled representation via extraction attention module and supervised contrastive learning for cross-domain sequential recommender. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 479–490.

[96] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. 2022. A survey on cross-domain recommendation: Taxonomies, methods, and future directions. *ACM Transactions on Information Systems* 41, 2 (2022), 1–39.

[97] He Zhang, Ying Sun, Weiyu Guo, Yafei Liu, Haonan Lu, Xiaodong Lin, and Hui Xiong. 2023. Interactive interior design recommendation via coarse-to-fine multimodal reinforcement learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, 6472–6480.

[98] Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. 2022. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4741–4753.

[99] Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, et al. 2023. Siren's song in the AI Ocean: A survey on hallucination in large language models. arXiv:2309.01219. Retrieved from https://doi.org/10.48550/arXiv.2309.01219

[100] Yuting Zhang, Ying Sun, Fuzhen Zhuang, Yongchun Zhu, Zhulin An, and Yongjun Xu. 2023. Triple dual learning for opinion-based explainable recommendation. *ACM Transactions on Information Systems* 42, 3 (2023), 1–27.
[101] Yuting Zhang, Yiqing Wu, Ruidong Han, Ying Sun, Yongchun Zhu, Xiang Li, Wei Lin, Fuzhen Zhuang, Zhulin An, and Yongjun Xu. 2024. Unified dual-intent translation for joint modeling of search and recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6291–6300.
[102] Zizhuo Zhang and Bang Wang. 2023. Prompt learning for news recommendation. In *SIGIR '23: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval,* 227–237. DOI: https://doi.org/10.1145/3539618.3591752
[103] Anhao Zhao, Fanghua Ye, Jinlan Fu, and Xiaoyu Shen. 2024. Unveiling in-context learning: a coordinate system to understand its working mechanism. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. ACM, 12375–12400. DOI: https://doi.org/10.18653/v1/2024.emnlp-main.689
[104] Chuang Zhao, Hongke Zhao, Xiaomeng Li, Ming He, Jiahui Wang, and Jianping Fan. 2023. Cross-domain recommendation via progressive structural alignment. *IEEE Transactions on Knowledge and Data Engineering* 36, 6 (2023), 2401–2415.
[105] Xiaoyun Zhao, Ning Yang, and Philip S Yu. 2022. Multi-sparse-domain collaborative recommendation via enhanced comprehensive aspect preference learning. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 1452–1460.
[106] Yongchun Zhu, Kaikai Ge, Fuzhen Zhuang, Ruobing Xie, Dongbo Xi, Xu Zhang, Leyu Lin, and Qing He. 2021. Transfer-meta framework for cross-domain recommendation to cold-start users. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1813–1817.
[107] Yongchun Zhu, Zhenwei Tang, Yudan Liu, Fuzhen Zhuang, Ruobing Xie, Xu Zhang, Leyu Lin, and Qing He. 2022. Personalized transfer of user preferences for cross-domain recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 1507–1515.

Appendix

A.1 Full Version of The Designed Prompt

We present the meta format of the prompt for candidate-free cross-domain interaction generation:

```
  ### Instruction:
  In the

source domain

, there are various products that a user has purchased.
Based on these, recommend some

target domain

products the user might have an
interest in purchasing. Some examples that demonstrate the input of

source
domain

product history and the output of recommended

target domain

products are given below. Do not simply select items from the examples.
```

```
  ### Example 1
  Input:[

example source item 1_1

,

example source item 1_2

, ...]
  Output:
  -

example target item 1_1


  -

example target item 1_2


  - ...
  ### Example ...
  ...


  ### Input: [

input source item 1

,

input source item 2

, ...]
```

A.2 Experiments on More Datasets

We additionally constructed a CDSR dataset, **Cloth-Beauty**, on top of the Amazon datasets. The data processing followed a similar approach to that used for the Movie-Book and Food-Kitchen datasets. Particularly, we applied the 5-core setting to filter out inactive users and items. Moreover, we segmented each user's interactions into smaller sequences, using a 6-month interval as the cutoff. The statistics for this dataset are shown in Table A1. We then evaluate the overall CDSR performance of LLMCDSR against baseline models on the Cloth-Beauty dataset. As shown in Table A2, the results are consistent with those in Section 5.3, demonstrating the effectiveness of LLMCDSR across a broader range of recommendation scenarios.

Table A1. Cloth-Beauty Dataset Statistics

Dataset		Cloth-Beauty Cloth	Beauty
# Users	Non-overlapped	9,895	6,865
	Overlapped	2,860	
# Interaction records	Single-domain	10,292	7,573
	Cross-domain	3,387	
# Items		21,368	11,728
Average sequence length		9.02	
Density		0.0129%	

Table A2. The Overall Recommendation Performance on the Cloth-Beauty Scenario

	Cloth						Beauty					
	NDCG			HR			NDCG			HR		
	@3	@5	@10	@3	@5	@10	@3	@5	@10	@3	@5	@10
TPUF	0.0092	0.0092	0.0182	0.0183	0.0183	0.0459	0.0139	0.0175	0.0249	0.0222	0.0311	0.0533
PiNet	0.0166	0.0166	0.0245	0.0263	0.0263	0.0526	0.0093	0.0093	0.0123	0.0185	0.0185	0.0278
C2DSR	0.0263	0.0263	0.0263	0.0526	0.0526	0.0526	0.0046	0.0086	0.0086	0.0093	0.0185	0.0185
MGCL	0.0132	0.0132	0.0132	0.0263	0.0263	0.0263	0.0000	0.0036	0.0125	0.0000	0.0093	0.0370
FPMC	0.0116	0.0155	0.0216	0.0183	0.0275	0.0459	0.0412	0.0505	0.0656	0.0489	0.0711	0.1200
Caser	0.0195	0.0195	0.0251	0.0275	0.0275	0.0459	0.0117	0.0134	0.0206	0.0178	0.0222	0.0444
FEARec	0.0136	0.0185	0.0199	0.0203	0.0248	0.0397	0.0211	0.0281	0.0368	0.0297	0.0314	0.0382
GRU4Rec	0.0275	0.0311	0.0341	0.0275	0.0367	0.0459	0.0072	0.0166	0.0263	0.0089	0.0311	0.0622
SRGNN	0.0162	0.0237	0.0291	0.0275	0.0459	0.0642	0.0189	0.0262	0.0302	0.0222	0.0400	0.0533
SASRec	0.0173	0.0198	0.0276	0.0245	0.0306	0.0551	0.0449	0.0600	0.0731	0.0563	0.0933	0.1333
LLMCDSR	**0.0391**	**0.0506**	**0.0594**	**0.0642**	**0.0917**	**0.1193**	**0.0586**	**0.0714**	**0.0854**	**0.0844**	**0.1156**	**0.1600**

The best results are highlighted in bold, and the second-best results are underlined.

Received 2 February 2024; revised 12 November 2024; accepted 2 January 2025