Retrieval Augmented Cross-Domain LifeLong Behavior Modeling for Enhancing Click-through Rate Prediction

Xing Tang*
Shenzhen Technology University
Shenzhen, China

Chaohua Yang*
Shenzhen University
Shenzhen, China

Yuwen Fu*
FiT, Tencent
Shenzhen, China

Dongyang Ao
FiT, Tencent
Shenzhen, China

Shiwei Li†
Huazhong University of Science and Technology
Wuhan, China

Fuyuan Lyu
McGill University
Montreal, Canada

Dugang Liu
Shenzhen University
Shenzhen, China

Xiuqiang He†
Shenzhen Technology University
Shenzhen, China

Abstract

Lifelong behavior modeling for single-domain has been widely investigated in industry click-through (CTR) prediction. However, some domains do not always have rich historical behaviors in online platforms, so cross-domain lifelong behavior modeling is overlooked. This paper proposes a novel retrieval augmented lifelong cross-domain net (RAL-CDNet) to address the challenges in cross-domain lifelong behavior modeling. There are three components in RAL-CDNet, i.e., cross-domain retrieval unit, cross-domain alignment unit, and cross-net. As the general search unit in the previous study, a cross-domain retrieval unit features a retrieval augmented paradigm that utilizes a pre-trained language model to learn the intrinsic textual information of user behaviors and generates the sequential behaviors from the source domain based on sequential behaviors in the target domain. The retrieval augmented behaviors can achieve consistency and capture accurate hidden interest for target domain CTR prediction. Furthermore, we propose the cross-domain alignment unit to align the embeddings across domains by adding a semantic-guided contrastive loss and auxiliary task loss in the source domain. This allows the embeddings to be consistent across domains and have enough source information to capture the cross-domain relation. Finally, the cross-net utilizes two-level attention techniques to enhance the final prediction in the target domain. We conduct extensive experiments on both a public dataset and an industrial dataset from the WeChat advertising platform to demonstrate the effectiveness of RAL-CDNet in terms of offline and online metrics.

CCS Concepts

- **Information systems → Recommender systems.**

Keywords

Click-through Rate Prediction, Cross-domain , LifeLong User Behavior

ACM Reference Format:
Xing Tang, Chaohua Yang, Yuwen Fu, Dongyang Ao, Shiwei Li, Fuyuan Lyu, Dugang Liu, and Xiuqiang He. 2025. Retrieval Augmented Cross-Domain LifeLong Behavior Modeling for Enhancing Click-through Rate Prediction . In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2 (KDD '25), August 3–7, 2025, Toronto, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3711896.3737261

1 Introduction

Click-through rate (CTR) prediction has played a crucial role in online business platforms. It aims to predict the probability of the user clicking the candidate items (e.g., advertisement, video) [4, 23, 32]. The industrial CTR model input mainly consists of user and item information, which provides personalization. To fully utilize user information, lifelong user behavior modeling has been widely used in extracting a user's hidden interests concerning the candidate item, which acts as a key component in CTR prediction [26, 29, 38].

Nevertheless, the sequential behaviors in some small-scale domains are sparse, posing a challenge to capturing the user's hidden interest. As an example illustrated in Fig. 1a, users click on some advertising coupons related to APPs in the target domain, which is sparse and makes it hard to reflect the user's interest accurately. Moreover, users will also spend a lot on corresponding APPs in the source domain, resulting in lifelong behaviors. It can be noticed that the length of behaviors in the source domain is nearly eight times more than in the target domain from Fig. 1b, which means that the lifelong user behaviors in the source domain help extract hidden interests more accurately. Therefore, how to transfer the lifelong behavior information from the source domain and make comprehensive predictions in the target domain is essential. A promising solution is the cross-domain CTR model [18, 24], which aims to

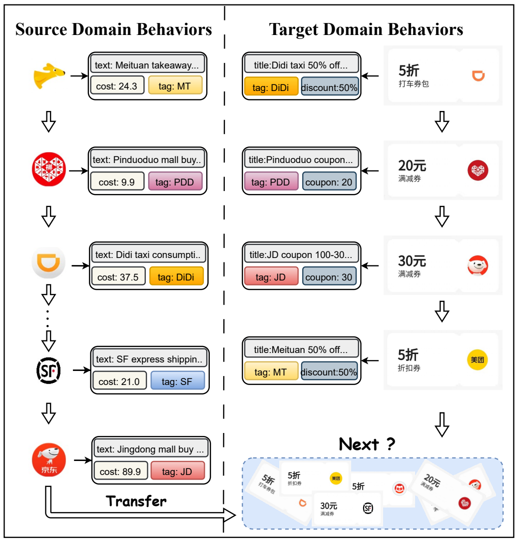
---

*Both authors contributed equally to this research.
†Corresponding Authors.

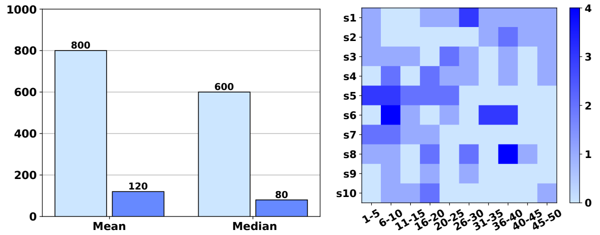
KDD '25, Toronto, ON, Canada

ACM ISBN 979-8-4007-1454-2/2025/08
https://doi.org/10.1145/3711896.3737261

leverage auxiliary and rich user lifelong behavior from a source domain to improve the performance of the target domain.



(a) The example from WeChat advertising platform



(b) Statistics of the length of behavior sequence (c) The relevant items in top-k items

Figure 1: (a) An example in the WeChat advertising platform.(b) A comparison between the statistics of the length of behavior sequences of the source domain and the target domain. (c) A map illustrated the distribution of relevant items in the set of embedding-based top 50 items for 10 users.

Generally, lifelong user behavior modeling in the single-domain mostly follows the two cascading stages [27, 28]: the General Search Unit (GSU) handles lifelong user behaviors coarsely to reduce the behaviors relevant to the target candidate item, and the Exact Search Unit (ESU) extracts the interest representation from the output behaviors of GSU to help make the final prediction. However, adopting the lifelong behavior modeling paradigm directly to the cross-domain CTR model will face three particular challenges. First, GSU will face *inconsistency* in cross-domain lifelong user behavior modeling. GSU usually utilizes either expressive item embeddings [3, 5] or item intrinsic features to identify relevant items in lifelong behaviors [28, 29]. On the one hand, item embeddings in the source domain are supervised by the label from the target domain. On the other hand, the features are not similar across both domains. How to sift through lifelong user behaviors effectively determines which relevant items from the source domain can be transferred to the target domain. And cross-domain modeling significantly increases the difficulty and complexity of searching in GSU. Second, the way to search for the behaviors in GSU relevant to the target candidate item will suffer from *sparse* issue. Specifically, previous studies will take the candidate item or behavior as the query to search for the relevant behaviors [2]. However, the cross-domain sequence contains more noise, leading to fewer relevant items in the source domain. As Fig. 1c shows, the set of embedding-based top-k similar items only recalls a few relevant items in the source domain. The sparse sequence will then make it hard to extract user interest. Third, ESU must deal with *alignment* between the source domain and the target domain. In cross-domain modeling, the only supervised signal is from the target domain, which restricts the optimization of the effective source domain embedding and makes it difficult to capture the cross-domain relation, thus degrading the performance [15].

This paper addresses these challenges by proposing a novel retrieval augmented lifelong cross-domain net (**RAL-CDNet**) for cross-domain lifelong behavior modeling. To tackle the inconsistency issue, we introduce a cross-domain retrieval unit (CD-RU), which is inspired by retrieval augmented generation incorporating extern knowledge in the large language model [9, 10]. The intrinsic textual description of the item will be transformed into an embedding representation by a pre-trained language model (PLM) [14], which then serves as a trace to determine the relevant items in our CD-RU. Notice that the textual information is intrinsic across domains and contains less noise, thus helping more accurately identify the relevant behaviors. Moreover, we input not only candidate behavior but all the sequential behaviors into the CD-RU. Then, the CD-RU generates new sequential behaviors from the source domain. In other words, we utilize CD-RU to retrieve source domain sequential behaviors according to the sequence in the target domain, which mitigates the sparse issue. We propose a cross-domain alignment unit (CD-AU) incorporating a semantically guided contrastive loss and source domain auxiliary task to align the source and target embeddings. Finally, we also design a cross-net with a two-level attention mechanism to learn the two sequential behaviors enhancing the CTR prediction [34]. Our contributions are summarized as follows:

- This paper first identifies related challenges in cross-domain industry scenarios for lifelong user behavior modeling. Then, a novel framework named **RAL-CDNet** is proposed. By addressing the identified challenges, RAL-CDNet can effectively transfer lifelong user behaviors from the source domain to the target domain.
- We propose CD-RU, CD-AU, and cross-net, consisting of RAL-CDNet. In CD-RU, we adopt a retrieval-augmented paradigm to generate the source domain sequential behavior according to the target domain sequence. The CD-AU aligns the representation across the domain and then cross-net utilizes the transferred information to enhance the CTR prediction in the target domain.
- Extensive offline experiments on public and real-world product datasets are conducted. The results demonstrate the superiority of the RAL-CDNet. Furthermore, we also deploy our RAL-CDNet on our large-scale advertising platform, and notable online A/B testing improvements verify the effectiveness of the RAL-CDNet in enhancing the prediction.

2 Related work

In this section, we briefly review related work on two topics: cross-domain CTR prediction and lifelong user behavior modeling.

2.1 Cross-domain CTR prediction

Several arrays of studies contribute to cross-domain recommendation [16, 25, 39]. The most related topic to our work is cross-domain CTR prediction, which mainly focuses on leveraging source domain information to improve the performance of the target domain [6, 18, 22, 24]. MiNet [24] considers three types of user interest in cross-domain CTR prediction and designs two levels of attention to fuse the information inter-sequence and intra-sequence. To further fuse the information, DASL [18] introduces dual embedding and dual attention. CDANet [6] proposes a cross-domain augmentation network that can perform explicit knowledge transfer between two domains, which is more flexible for transferring information across domains. Our work differs from these in that it focuses on leveraging lifelong user behavior from the source domain to help with CTR prediction in the target domain.

2.2 Lifelong User Behavior Modeling

The modeling of rich user behaviors has been a key component for improving the performance in the industry CTR prediction [26, 30, 37, 38]. With the exponential growth of data, the computational demand for lifelong user behaviors becomes too heavy to tackle within one model. The SIM [27] and UBR4CTR [28, 29] introduce the two-stage cascaded framework to divide the process into a General Search Unit and an Exact Search Unit. On the one hand, some works are devoted to designing various relevant metrics and ways to find the relevant items. For example, ETA [5] uses the locality-sensitive hash to encode item embeddings trained by ESU and retrieves relevant behaviors from long-term behaviors via Hamming distance. On the other hand, some works address the discrepancy between GSU and ESU. TWIN [3, 33] designs an identical target-behavior relevance metric for both stages to maintain end-to-end training. Notice that extending these works to cross-domain lifelong user behavior modeling is nontrivial and challenging, and may degrade the performance due to the transfer of information. LCN [15] first tackles the challenge by introducing a lifelong attention pyramid. However, the inconsistent embeddings across domains and the noise present in lifelong user behaviors from the source domain still leave room for performance improvement.

3 Preliminary

We formalize the modeling problem in this section. First, we consider the features of a user u in both the target and source domains consisting of three categories:

- Basic user features $\{\mathbf{x}_u\}$.
- User short-term behaviors in target domain t, represented by $\mathcal{T}_u = \{v_1^t, \cdots, v_j^t, \cdots, v_{z_u}^t\}$.
- User lifelong behaviors in the source domain s,represented by $\mathcal{S}_u^l = \{v_1^s, \cdots, v_i^s, \cdots, v_{m_u}^s\}$.

Notice that the length of the sequence $\mathcal{T}_u$ is z_u, which is far smaller than the length m_u of the lifelong sequence $\mathcal{S}_u^l$. The CTR prediction in the target domain aims to estimate the click probability of user-item pair $\langle u, v\rangle$ with respect to item features $\{\mathbf{x}_v\}$ as follows:

$$p_{uv} = \mathcal{F}_\theta(\mathbf{x}_u, \mathbf{x}_v \mid \mathcal{T}_u), \tag{1}$$

where θ represents the parameters of the prediction model $\mathcal{F}$. With the aid of transferring lifelong user behaviors from the source domain, we further introduce the retrieval function $\mathcal{R}$, and $\mathcal{R}(\mathcal{T}_u, \mathcal{S}_u^l)$ with short-term behaviors in the target domain and lifelong behavior in source domain as input. The cross-domain lifelong behavior modeling is formulated as:

$$p_{uv} = \mathcal{F}_\theta(\mathbf{x}_u, \mathbf{x}_v \mid \mathcal{T}_u, \mathcal{R}(\mathcal{T}_u, \mathcal{S}_u^l)). \tag{2}$$

Our RAL-CDNet appropriately aims to model the $\mathcal{F}$ and $\mathcal{R}$. The optimized cross-entropy loss function is defined as follows:

$$\mathcal{L}_{CTR} = -\frac{1}{B}\sum_{i=1}^{B}(y_i \log(p_{u_i,v_i}) + (1-y_i)\log(1-p_{u_i,v_i})), \tag{3}$$

where $y_i \in \{0, 1\}$ is the label that indicates whether the user clicks or not, and B denotes the total number of user-item sample pairs in a batch.

4 Methodology

In this section, we elaborate on the proposed method in detail. The overall framework of RAL-CDNet is illustrated in Fig. 3, which mainly consists of three components: cross-domain retrieval unit, cross-domain alignment unit, and cross-net, respectively. The RAL-CDNet's inputs include user features, candidate item features, and behaviors in both the target and source domains. First, it utilizes a cross-domain retrieval unit to generate relevant short-term sequences from the source domain. After the embedding layer, two short-term sequential behaviors are aligned with a cross-domain alignment unit. Meanwhile, the following cross-net will give the final click-through prediction.

4.1 Cross-Domain Retrieval Unit

The cross-domain retrieval unit generates relevant behaviors in the source domain according to the short-term behaviors in the target domain. The associated textual information related to lifelong user behaviors in the source domain can be denoted as $\{t_1^s, \cdots, t_{m_u}^s\}$. Similarly, the sequence of short-term behaviors in the target domain is $\{t_1^t, \cdots, t_{z_u}^t\}$. Note that previous works have investigated that semantic knowledge from textual CTR data can be a complementary element for improving performance [8, 11, 20]. In our RAL-CDNet, the textual data contains the intrinsic information that connects the source and target domain, as an example shown in Fig. 2.

Text t_i^s of v_i^s : "Pinduoduo Mall purchased a piece of clothing and paid 120 RMB."

Text t_j^t of v_j^t : "Pinduoduo coupon: Spend 20 RMB and get 10 RMB off."

Figure 2: An example of textual data for source domain and target domain item.

Specifically, we compute the textual embedding related to the items in the target domain, i.e., for any item v_j^t in the target domain,

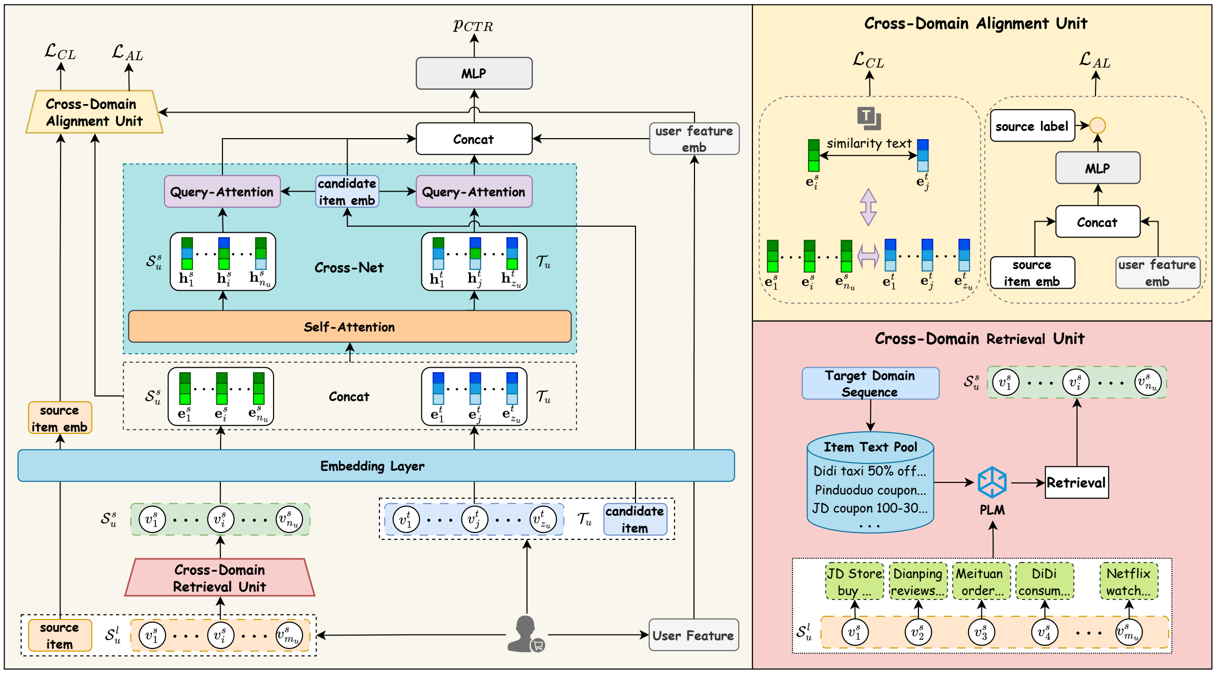


Figure 3: The overall framework of our retrieval augmented lifelong cross-domain net (RAL-CDNet). The model has three components: (1) the bottom right illustrates the cross-domain retrieval unit, which retrieves sequences from the lifelong user behavior in the source domain. (2) the upright shows the cross-domain alignment unit aligns the embeddings across domains and explores the relation between domains. (3) The blue part on the left is a cross-net, fusing information for the final prediction.

we get the textual embedding $\mathbf{rep}_j^t$ with a PLM. We also use the same PLM to calculate the textual representation $\mathbf{rep}_i^s$ for lifelong user behaviors. To sift through the lifelong user behaviors, we introduce a mask vector $\mathbf{m} = \{0,1\}^{m_u}$, the mask value m_i is then defined as follows:

$$m_i = 1 \iff \exists j \in \{1,\cdots,z_u\}, \cos(\mathbf{rep}_i^s, \mathbf{rep}_j^t) > \theta,$$
$$\text{where} \quad \cos(\mathbf{rep}_i^s, \mathbf{rep}^t) = \frac{\mathbf{rep}_i^s \cdot \mathbf{rep}_j^t}{\|\mathbf{rep}_i^s\|\|\mathbf{rep}_j^t\|}. \tag{4}$$

Where the θ is a threshold, which is determined as a hyper-parameter, and with the mask vector, we keep the lifelong behaviors only related to the input short-term sequence in the target domain.

The retrieval function $\mathcal{R}$ in Eq. 2 can be further defined as:

$$\mathcal{R} = \mathbf{m} \cdot \mathcal{S}_u^l. \tag{5}$$

We drop the zero value element in the output, and the output of $\mathcal{R}$ is $\mathcal{S}_u^s$ with the length of the sequence n_u, which is smaller than m_u. Notice that we pre-compute the similarity scores in our online version, and we can fetch the score according to the item pair, accelerating the computation.

4.2 Cross-Domain Alignment Unit

The retrieved sequence $\mathcal{S}_u^s$ and the sequence $\mathcal{T}_u$ in the target domain will be embedded into the latent vectors. However, both the embeddings are only optimized following the target supervised labels, making it hard to capture the relation between the source and target domains. Moreover, a gap exists between the embedding for the source domain and the target domain because the retrieved sequence is based on the behaviors from the source domain.

To overcome these challenges, we propose the cross-domain alignment unit, which is responsible for aligning the embeddings across domains. To preserve the information from the source domain, we simply introduce an auxiliary loss to ensure that the embedding of the source behaviors can be used to predict the click probability in the source domain. For the user u_i, the source item to be predicted can be denoted as v_i^s. With the aid of the embedding layer, both features are converted to embedding, $\mathbf{e}_{u_i}$ and $\mathbf{e}_{v_i^s}$. Our auxiliary task is to predict whether source item v_i^s can be clicked:

$$p_{click} = \sigma(MLP(\mathbf{e}_{u_i}, \mathbf{e}_{v_i^s})), \tag{6}$$

where $\sigma(x) = \frac{1}{1+\exp(-x)}$, and MLP is a multi-layer perception with activation function. Furthermore, the loss function is:

$$\mathcal{L}_{SL} = -[y_i \cdot \log(p_{click}) + (1-y_i) \cdot \log(1-p_{click})], \tag{7}$$

where $y_i \in \{0,1\}$ is the click label of the user-item pair, indicating whether the item will be clicked or not in the source domain.

The auxiliary task only guarantees that the information from the source domain is preserved. We also need an extra loss to align the embeddings and explore the relation across domains. We construct a contrastive loss based on the textual similarity [1, 35, 36]. To begin with, we select an item embedding $\mathbf{e}_i^s$ from the $\mathcal{S}_u^s$, the similarity score in the cross-domain retrieval unit is re-used to find a textual similar item in $\mathcal{T}_u$, and the embedding of this item is denoted as $\mathbf{e}^t_j$, which constitutes the positive sample pair $\langle v^s_i, v^t_j \rangle$. Regarding the negative pairs, we employ a parallel approach to the positive pair, resulting in pairs with similarity scores below the threshold. The contrastive loss is summarized as follows:

$$\begin{aligned}\mathcal{L}_{CL} &= -\log \frac{\sum_i^{n_u} \sum_j^{z_u} m_{ij} \cdot \exp((\mathbf{e}^{\mathrm{s}}_{\mathrm{i}} \odot \mathbf{e}^{\mathrm{t}}_{\mathrm{j}})/\tau)}{\sum_i^{n_u} \sum_j^{z_u} (1 - m_{ij}) \cdot \exp((\mathbf{e}^{\mathrm{s}}_{\mathrm{i}} \odot \mathbf{e}^{\mathrm{t}}_{\mathrm{j}})/\tau)} \\ \text{s.t.} \quad m_{ij} &= \mathbb{I}(\cos(\mathbf{rep}^s_i, \mathbf{rep}^t_j) > \theta), \end{aligned} \tag{8}$$

where $\odot$ denotes the vector inner product operation, and τ is the adjusted temperature coefficient.

4.3 Cross-Net

The cross-net gives the final prediction of CTR in the target domain. The major challenge for cross-net here is how to fuse the information between the two sequences $\mathcal{S}^s_u$ and $\mathcal{T}_u$. We design a two-level attention mechanism to fuse the information here.

Due to the previous alignment mechanism, we first concat both two embeddings of sequences, $\mathbf{b}_u = \{\mathbf{e}^s_1, \cdots, \mathbf{e}^s_{n_u}; \mathbf{e}^t_1, \cdots, \mathbf{e}^t_{z_u}\}$. The attention process can be used with the following formulas:

$$\begin{aligned} &\mathbf{r}_u = softmax(\frac{QK^T}{\sqrt{d}}), \mathbf{h}_u = \mathbf{r}_u \cdot \mathbf{b}_u \\ &\text{where} \quad Q = W^Q \mathbf{b}_u, K = W^K \mathbf{b}_u \end{aligned} \tag{9}$$

where W^Q and W^K denote the attention weights, d represents the inner dimension, the $\mathbf{r}_u$ is the attention scores associated with the $\mathbf{b}_u$, and the $\mathbf{h}_u$ is the weighted sequence. Note that we conduct the attention operation on the whole sequence $\mathbf{b}_u$ without discriminating between $\mathcal{S}^s_u$ and $\mathcal{T}_u$, which leads to fusing the information across two domains.

We further introduce query attention to associate the embedding with the candidate item $\mathbf{e}_v$. Because different sequences contribute differently to the final prediction, we divide the $\mathbf{h}_u$ into two parts related to two domains again, i.e., $\mathbf{h}^s_u = \{\mathbf{h}^s_1, \cdots, \mathbf{h}^s_{n_u}\}$ and $\mathbf{h}^t_u = \{\mathbf{h}^t_1, \cdots, \mathbf{h}^t_{z_u}\}$ respectively. The query attention process is performed on each sequence, and we take the process on the source sequence as an example:

$$\begin{aligned} &\mathbf{e}^s_u = softmax(\frac{Q_q K_q^T}{\sqrt{d}}) V_q, \\ &\text{where} \quad Q_q = W_q^Q \mathbf{e}_v, K_q = W_q^K \mathbf{h}^s_u, V_q = W_q^V \mathbf{h}^s_u. \end{aligned} \tag{10}$$

This attention process differs from the first level of attention only in the query, which is the candidate item.

Combined with the sequential presentation, user representation, and item representation, the final prediction in Eq. 2 can be reformulated:

$$p_{uv} = \mathcal{F}_\theta(\mathbf{e}_u, \mathbf{e}_v, \mathbf{e}^s_u, \mathbf{e}^t_u) \tag{11}$$

The $\mathcal{F}$ can be a single MLP or other model, such as DeepFM [13]. The final optimized loss function is a combination of three losses:

$$\mathcal{L} = \mathcal{L}_{CTR} + \lambda_{CL} \mathcal{L}_{CL} + \lambda_{AL} \mathcal{L}_{AL}, \tag{12}$$

where λ_{CL} and λ_{AL} denotes the factor to control the importance of $\mathcal{L}_{CL}$ and $\mathcal{L}_{AL}$ respectively.

5 Experiments

In this section, we conduct experiments to answer the following five key questions, demonstrating the superiority of our RAL-CDNet.

- RQ1: How does RAL-CDNet perform compared to the baselines?
- RQ2: What is the role of some key components in RAL-CDNet?
- RQ3: What is the impact of different sequence lengths and pre-trained models on the performance of RAL-CDNet?
- RQ4: How effective is the CD-RU in RAL-CDNet?
- RQ5: How does RAL-CDNet perform in online business scenarios?

5.1 Experiment Setup

5.1.1 Datasets. We conducted experiments using two datasets: one is the public Amazon dataset, and the other is an industrial dataset. The processed datasets' statistics are summarized in Table 1.

Amazon Dataset[1]. The Amazon dataset is extensively utilized to assess the performance of recommendation systems. In this study, we selected three of the most commonly used categories, i.e., Books, Movies & TV, and CDs & Vinyl, for cross-domain CTR prediction tasks. Specifically, we treated Book as the source domain, while Movie and CD were used as two target domains. During preprocessing, we only retain users who co-occur in both the source and target domains and have at least 5 ratings for items with metadata. We converted ratings of 4 or 5 to label 1 and others to label 0. To simulate the CTR prediction of real industrial recommendation scenarios (i.e., predict the future CTR), we sorted user logs chronologically. For each user, the most recent rating was used to form the test set, the second most recent rating was used for the validation set, and the remaining ratings constituted the training set.

Industrial Dataset. This dataset is collected from the WeChat advertising platform's traffic logs and contains user behavior sequences related to successful coupon recharges and WeChat payments. We collected each user's 30-day behavior sequence for coupon items and 365-day behavior sequence for payment items (the maximum length is 3,000). The payment item sequence is specified as the sequence in the source domain. The label of each sample is the user's click operation on the coupon item. The dataset contains 500 million records from 200 million users, collected over 31 days. We divided the dataset based on time, using data from the first 30 days for training and data from the 31st day for testing.

5.1.2 Metrics. For offline evaluation, following previous work [19, 21], we utilized two commonly used metrics for CTR prediction tasks: the Area Under the ROC Curve (AUC) and the logarithmic loss (logloss). For online A/B testing, the evaluation was conducted using two key performance indicators: CTR and Cost Per Mille (CPM), measuring advertising revenue per thousand impressions.

5.1.3 Baselines Methods. To comprehensively evaluate our proposed RAL-CDNet, we selected a series of methods from multiple directions, including single-domain, single-domain sequential, cross-domain, and cross-domain sequential methods, for comparison. Below is an overview of the baseline methods:

- **DNN** [7]. It is a Deep Neural Network consisting of an embedding layer, multiple FC layers, and an output layer.

---

[1] https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

Table 1: Statistics of the processed datasets.

Dataset	Amazon				Industrial	
	Source:Book	Target:Movie	Source:Book	Target:CD	Source	Target
#Shared users	133,103		54,835		272,415,397	
#Items	1,006,743	121,386	691,947	238,329	9,572	1,033
#Train instances	1,682,884		852,743		501,349,887	
#Valid instances	133,103		54,835		18,617,498	
#Test instances	133,103		54,835		18,617,498	
Max.Len/Avg.Len #clicked	9,463/24.12	1,919/11.58	9,463/28.16	2,938/15.18	499,284/887.68	86/15.18

- **DeepFM** [13]. It is a recommendation model that combines Factorization Machines (FM) [31] and Deep Neural Networks (DNN) to model both low- and high-order feature interactions.
- **DIN** [38]. The Deep Interest Network (DIN) models users' dynamic interests based on their historical behaviors to enhance the accuracy of CTR prediction.
- **SIM** [27]. This early work performs lifelong sequence modeling through the GSU and ESU stages while designing SIM(Hard) and SIM(Soft) based on different search strategies in the GSU stage.
- **TWIN** [3]. This is a long sequence modeling method that enhances the consistency between GSU and ESU by applying cross-feature dimension compression.
- **CoNet** [16]. This is a classic method for cross-domain recommendation, which enables cross-domain knowledge transfer through a cross-connection.
- **CDAnet** [6]. This approach consists of a translation network and an augmentation network designed to perform explicit knowledge transfer between two domains.
- **MiNet** [24]. Mixed Interest Network (MiNet) simultaneously models three types of user interests: long-term interests across domains, short-term interests from the source domain, and short-term interests in the target domain.
- **LCN** [15]. This is a Lifelong Cross Network (LCN) for cross-domain lifelong sequence modeling, which mainly consists of two components: the Cross Representation Production module and the Lifelong Attention Pyramid module.
- **RAL-CDNet(Hard)**. This is a simplified version of our proposed RAL-CDNet, which performs cross-domain retrieval through hard matching of textual keywords.

5.1.4 Parameter Settings. For general hyperparameters, all features, including items and user IDs, are mapped to distinct embedding spaces with a uniform embedding size of 16. The model parameters are initialized using the Xavier initialization method [12], and the model is optimized using the Adam optimizer [17]. We select the optimal learning rate from {1e-3, 5e-4, 1e-4, 5e-5, 1e-5} and the l_2 regularization from {1e-3, 5e-4, 1e-4, 5e-5, 1e-5, 5e-6, 1e-6}. We set the batch size to 512 and 4096 for the public and industrial datasets, respectively. For the hyperparameters of RAL-CDNet, we select the optimal threshold θ from $\{0.5, 0.6, 0.7, 0.8\}$, and the optimal weight coefficient λ_{CL} and λ_{AL} from {1e-1, 5e-2, 1e-2, 5e-3, 1e-3, 5e-4, 1e-4} and {0.0, 0.1, 0.2, 0.3}, respectively. Our model is developed using TensorFlow. For baseline methods, we utilize the open source implementations for DNN [7], DeepFM [13] and DIN [38], CoNet [16] and MiNet [24]. Since there are no existing implementations available for SIM [27], TWIN [3], CDAnet [6], and LCN [15], we re-implement them using the information provided in the original paper. Note that all baseline models are tuned within the same hyperparameter setting range to achieve optimal performance.

5.2 RQ1: Overall Performance

In this subsection, we compare RAL-CDNet with four types of baselines: single-domain, single-domain sequential, cross-domain, and cross-domain sequential recommendation models. The overall performance of RAL-CDNet and the baseline methods is summarized in Table 2.

From Table 2, we can make the following observations: 1) On both datasets, single-domain sequential methods (e.g., DIN, SIM, TWIN) and cross-domain methods (e.g., CoNet, CDAnet) generally outperform pure single-domain methods (e.g., DNN, DeepFM). This clearly indicates that incorporating user interaction sequence information and cross-domain information helps to more accurately model users' true preferences, while single-domain methods overlook these important factors. 2) Cross-domain methods (e.g., CoNet, CDAnet) significantly outperform single-domain sequential methods (e.g., DIN, SIM, TWIN) on the Amazon dataset. This may be because the user interaction behavior in the two target domains of the Amazon dataset is sparser, making cross-domain information more crucial for modeling user preferences in this case. 3) Cross-domain sequential methods (e.g., MiNet, LCN) outperform both single-domain sequential methods (e.g., DIN, SIM, TWIN) and cross-domain methods (e.g., CoNet, CDAnet) on both datasets. This indicates that incorporating sequential and cross-domain information is crucial for capturing user preferences in domains with sparse user interactions. 4) Our RAL-CDNet significantly outperforms all baseline methods on both datasets, demonstrating its effectiveness in cross-domain lifelong behavior modeling and its ability to enhance model performance. Furthermore, RAL-CDNet surpasses another cross-domain lifelong behavior modeling approach, LCN, further validating our key motivation: leveraging the intrinsic textual description of items and converting it into corresponding embedding representations for retrieval can help obtain more relevant items from the source domain, thereby enabling more accurate modeling of user preferences. Finally, RAL-CDNet outperforms RAL-CDNet(Hard), which indicates that converting

Table 2: Results on all datasets, where the best and second best results are marked in bold and underlined, respectively. Note that * indicates a significance level of $p \leq 0.05$ based on a two-sample t-test between our method and the best baseline.

Dataset	Amazon				Industrial	
Source → Target	Book → Movie		Book → CD		Payment → Ad	
	AUC↑	Logloss↓	AUC↑	Logloss↓	AUC↑	Logloss↓
DNN	0.7584	0.4413	0.7153	0.3255	0.8361	0.2100
DeepFM	0.7574	0.4421	0.7154	0.3306	0.8371	0.2096
DIN	0.7617	0.4672	0.7164	0.3293	0.8381	0.2091
SIM(Hard)	0.7612	0.5200	0.7158	0.3255	0.8380	0.2091
SIM(Soft)	0.7618	0.5128	0.7163	0.3253	0.8382	0.2091
TWIN	0.7624	0.4690	0.7176	0.3290	0.8379	0.2092
CoNet	0.7635	0.4270	0.7212	**0.3228**	0.8378	0.2093
CDAnet	0.7640	0.4267	0.7215	$\underline{0.3229}$	0.8381	0.2094
MiNet	0.7662	0.4311	0.7269	0.3449	0.8384	0.2092
LCN	0.7643	0.4318	0.7221	0.3323	0.8387	0.2089
RAL-CDNet(Hard)	$\underline{0.7708}$	$\underline{0.4230}$	$\underline{0.7301}$	0.3309	$\underline{0.8405}$	$\underline{0.2080}$
RAL-CDNet	**0.7717***	**0.4210***	**0.7310***	0.3291	**0.8407***	**0.2079**

textual descriptions into embeddings for retrieval is a crucial step in enhancing model performance.

5.3 RQ2: Ablation Study

In this subsection, we conduct an ablation study to assess the impact of key components in RAL-CDNet. Specifically, we first progressively remove key components from RAL-CDNet, including the cross-domain alignment unit (denoted as "**w/o** CD-AU"), the cross-domain retrieval unit (denoted as "**w/o** CD-RU"), and the cross-net (denoted as "**w/o** Cross-Net"). Then, we modify the functionality of the cross-domain retrieval unit in RAL-CDNet to retrieve the top-k relevant source domain items based on the candidate item embeddings, similar to LCN and train it in an end-to-end manner (denoted as "RAL-CDNet(E2E)"). The corresponding results are presented in Table 3. We can observe that removing any of the key components leads to performance degradation, indicating that the cross-domain alignment unit, cross-domain retrieval unit, and cross-net are all essential for our RAL-CDNet, with each contributing to cross-domain lifelong sequence modeling. Replacing the cross-domain retrieval unit with a method that retrieves top-k related source domain items based on the embedding similarity of candidate items also leads to performance degradation. This suggests that retrieving source domain items based on candidate item embedding similarity may result in fewer truly relevant items. Additionally, the end-to-end training approach increases the difficulty of embedding similarity retrieval.

Table 3: Ablation Analysis on our RAL-CDNet, where the best results are marked in bold.

Dataset	Amazon			
Source → Target	Book → Movie		Book → CD	
	AUC↑	Logloss↓	AUC↑	Logloss↓
w/o CD-AU	0.7686	0.4258	0.7272	0.3380
w/o CD-RU	0.7692	0.4292	0.7291	0.3384
w/o Cross-Net	0.7682	0.4275	0.7273	0.3303
RAL-CDNet(E2E)	0.7691	0.4261	0.7278	0.3384
RAL-CDNet	**0.7722**	**0.4209**	**0.7310**	**0.3291**

5.4 RQ3: In-depth Analysis of RAL-CDNet

Next, we provide more in-depth analysis and discussion on the impact of Sequence Length L and the PLM Size in RAL-CDNet. In addition, efficiency analysis and sensitivity analysis of parameter θ can be found in the appendix A

Impact of Sequence Length L. For sequence modeling methods, the sequence length L is a crucial parameter that directly affects the model performance. Our RAL-CDNet needs to consider the maximum sequence length settings for $\mathcal{S}_u^s$ and $\mathcal{T}_u$. For simplicity, we set the maximum sequence lengths of $\mathcal{S}_u^s$ and $\mathcal{T}_u$ to be the same. To analyze the impact of the maximum sequence length L on RAL-CDNet's performance, we conducted experiments with different settings of L, and the results are shown in Fig. 4. We can observe that as L increases, RAL-CDNet's performance improves on both Amazon datasets, indicating that the inclusion of more sequential behaviors is beneficial. However, when $L \geq 100$, the performance improvement becomes smaller, which may be due to the limited number of user interactions, making it increasingly difficult to introduce more sequential behaviors with larger L. Additionally, across different L settings, RAL-CDNet generally outperforms RAL-CDNet(Hard). It is worth noting that, to balance model training cost and performance, the final value of L for RAL-CDNet and other baseline methods is uniformly set to 50 for the experiments.

Impact of PLM Size. As mentioned earlier, we leverage a pre-trained language model (PLM) to convert the intrinsic textual descriptions of items into embedding representations for more efficient cross-domain retrieval. However, different PLMs interpret textual descriptions differently, leading to variations in embedding representations. These, in turn, affect cross-domain retrieval performance and ultimately influence the model's performance. To investigate the impact of different PLMs, we conducted experiments using three PLMs of varying sizes, with the results presented in

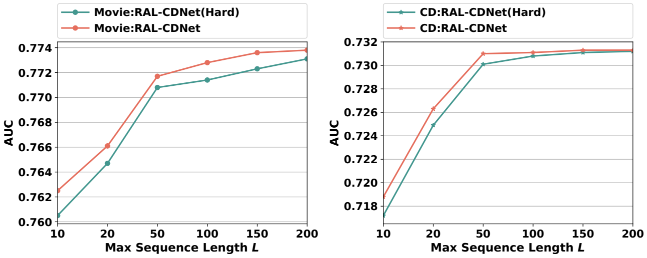


Figure 4: The impact of sequence length L on RAL-CDNet performance in the Amazon dataset.

Table 4: The Impact of Different Sizes of PLM, where the best results are marked in bold.

Dataset	Amazon			
Source ⟶ Target	Book ⟶ Movie		Book ⟶ CD	
PLM	AUC↑	Logloss↓	AUC↑	Logloss↓
BGE-Small (33M)	0.7695	0.4249	0.7304	0.3361
BGE-Base (109M)	0.7711	0.4305	0.7307	0.3352
BGE-M3 (568M)	**0.7717**	**0.4210**	**0.7310**	**0.3291**

Table 4. We can observe that PLMs with larger parameter scales perform better, which aligns with intuition. Larger PLMs generally have a stronger ability to understand textual descriptions and thus generate higher-quality embedding representations. Therefore, in the actual experiment, we ultimately chose BGE-M3[2] as the PLM for the cross-domain retrieval unit.

5.5 RQ4: Cross-Domain Retrieval Visualization

In this subsection, we demonstrate our method's cross-domain retrieval capability and effectiveness through several case studies. Specifically, we selected two cases from the Amazon dataset, corresponding to the hard match retrieval and our CD-RU retrieval, to briefly demonstrate the retrieval results, as shown in Fig. 5. As seen in Case 1, hard match retrieval is performed based on the textual description of the movie-domain item, "Action and Adventure," and the related items retrieved in the book domain also all belong to the "Action and Adventure" category. In contrast, since our CD-RU retrieval is based on the embedding representation of the item's textual description, it can capture a broader range of related items. Thus, in Case 2, we observe that items in the movie domain categorized as "Fantasy, Mystery, and Adventure" can retrieve the "Science Fiction" book, which essentially represents items aligned with the user's similar preferences. Overall, these two cases validate the effectiveness of our RAL-CDNet in cross-domain retrieval and demonstrate that CD-RU retrieval can discover a broader range of related items.

5.6 RQ5: Online Deployments

In this subsection, we report the online A/B testing results of our RAL-CDNet on the WeChat advertising platform over four consecutive weeks. The results further verify the effectiveness of our RAL-CDNet.

Figure 5: Case study of cross-domain retrieval visualization, where the source and target domains are the Amazon Book and Movie datasets, respectively.

Lifelong user behaviors from the source domain are payment behaviors collected through WeChat payments, and the target domain is advertising for some APP coupons during WeChat telephone recharging. We utilize payment behaviors to enhance the CTR prediction in coupon advertising. We collect user feedback over the A/B test period to calculate CTR and CPM metrics. Generally, compared with the baseline, the improvements are significant, with RAL-CDNet improving **+5.34%** in CTR and **+7.67%** in CPM. Moreover, we also give the details of online relative improvement ratios for four consecutive weeks in Fig. 6. It is easily concluded that our RAL-CDNet improves both CTR and CPM consistently.

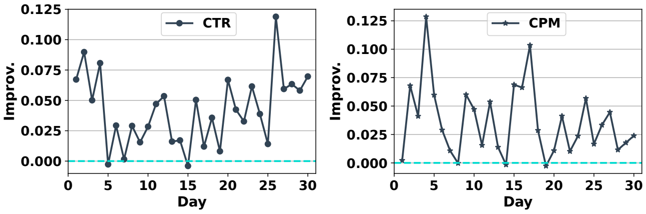


Figure 6: Online relative improvement ratios in consecutive four weeks(Left is the CTR improvement, Right is the CPM improvement).

6 Conclusion

In this paper, we investigate cross-domain lifelong behavior modeling to enhance CTR prediction. We propose a retrieval augmented lifelong cross-domain net (RAL-CDNet) composed of three components: a Cross-domain Retrieval Unit(CD-RU), a Cross-domain Alignment Unit(CD-AU), and a Cross-net. The CD-RU adopts a retrieval-augmented paradigm that utilizes textual CTR data to bridge the source and target domains. The CD-AU introduces two tasks to align embeddings across domains and explore the relation across domains. Finally, cross-net fuses the information together to give the prediction. Both online and offline experiments are conducted to verify the effectiveness of our RAL-CDNet. As for future work, we aim to introduce a large language model to optimize the two stages jointly.

---

[2] https://huggingface.co/BAAI/bge-m3

7 Acknowledgement

We thank the support of the SZTU University Research Project (No.20251061020002).

References

[1] 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 and Knowledge Management*. 138–147.

[2] Yue Cao, Xiaojiang Zhou, Jiaqi Feng, Peihao Huang, Yao Xiao, Dayao Chen, and Sheng Chen. 2022. Sampling Is all you need on modeling long-term user behaviors for CTR prediction. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*. 2974–2983.

[3] Jianxin Chang, Chenbin Zhang, Zhiyi Fu, Xiaoxue Zang, Lin Guan, Jing Lu, Yiqun Hui, Dewei Leng, Yanan Niu, Yang Song, and Kun Gai. 2023. TWIN: TWo-stage interest network for lifelong user behavior modeling in CTR prediction at kuaishou. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3785–3794.

[4] Olivier Chapelle, Eren Manavoglu, and Romer Rosales. 2014. Simple and scalable response prediction for display advertising. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 4 (2014), 1–34.

[5] Qiwei Chen, Changhua Pei, Shanshan Lv, Chao Li, Junfeng Ge, and Wenwu Ou. 2021. End-to-end user behavior retrieval in click-through rate prediction model. *arXiv preprint arXiv:2108.04468* (2021).

[6] Xu Chen, Zida Cheng, Jiangchao Yao, Chen Ju, Weilin Huang, Jinsong Lan, Xiaoyi Zeng, and Shuai Xiao. 2024. Enhancing cross-domain click-through rate prediction via explicit feature augmentation. In *Companion Proceedings of the ACM on Web Conference 2024*. 423–432.

[7] Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*. 7–10.

[8] Ziqiang Cui, Yunpeng Weng, Xing Tang, Fuyuan Lyu, Dugang Liu, Xiuqiang He, and Chen Ma. 2024. Comprehending knowledge graphs with large language models for recommender systems. *arXiv preprint arXiv:2410.12229* (2024).

[9] Ziqiang Cui, Yunpeng Weng, Xing Tang, Xiaokun Zhang, Dugang Liu, Shiwei Li, Peiyang Liu, Bowei He, Weihong Luo, Xiuqiang He, et al. 2025. Semantic retrieval augmented contrastive learning for sequential recommendation. *arXiv preprint arXiv:2503.04162* (2025).

[10] Wenqi Fan, Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, and Qing Li. 2024. A survey on rag meeting LLMs: Towards retrieval-augmented large language models. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 6491–6501.

[11] Binzong Geng, Zhaoxin Huan, Xiaolu Zhang, Yong He, Liang Zhang, Fajie Yuan, Jun Zhou, and Linjian Mo. 2024. Breaking the length barrier: LLM-enhanced CTR prediction in long textual user behaviors. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2311–2315.

[12] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*. 249–256.

[13] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: A factorization-machine based neural network for CTR prediction. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 1725–1731.

[14] Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, Wentao Han, Minlie Huang, Qin Jin, Yanyan Lan, Yang Liu, Zhiyuan Liu, Zhiwu Lu, Xipeng Qiu, Ruihua Song, Jie Tang, Ji-Rong Wen, Jinhui Yuan, Wayne Xin Zhao, and Jun Zhu. 2021. Pre-trained models: Past, present and future. *AI Open* 2 (2021), 225–250.

[15] Ruijie Hou, Zhaoyang Yang, Yu Ming, Hongyu Lu, Zhuobin Zheng, Yu Chen, Qinsong Zeng, and Ming Chen. 2024. Cross-domain lifelong sequential modeling for online click-through rate prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5116–5125.

[16] 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.

[17] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.

[18] Pan Li, Zhichao Jiang, Maofei Que, Yao Hu, and Alexander Tuzhilin. 2021. Dual attentive sequential learning for cross-domain click-through rate prediction. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3172–3180.

[19] Shiwei Li, Huifeng Guo, Lu Hou, Wei Zhang, Xing Tang, Ruiming Tang, Rui Zhang, and Ruixuan Li. 2023. Adaptive low-precision training for embeddings in click-through rate prediction. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. 4435–4443.

[20] Jianghao Lin, Bo Chen, Hangyu Wang, Yunjia Xi, Yanru Qu, Xinyi Dai, Kangning Zhang, Ruiming Tang, Yong Yu, and Weinan Zhang. 2024. ClickPrompt: CTR models are strong prompt generators for adapting language models to CTR prediction. In *Proceedings of the ACM Web Conference 2024*. 3319–3330.

[21] Dugang Liu, Chaohua Yang, Xing Tang, Yejing Wang, Fuyuan Lyu, Weihong Luo, Xiuqiang He, Zhong Ming, and Xiangyu Zhao. 2024. MultiFS: Automated multi-scenario feature selection in deep recommender systems. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 434–442.

[22] Lixin Liu, Yanling Wang, Tianming Wang, Dong Guan, Jiawei Wu, Jingxu Chen, Rong Xiao, Wenxiang Zhu, and Fei Fang. 2023. Continual transfer learning for cross-domain click-through rate prediction at Taobao. In *Companion Proceedings of the ACM Web Conference 2023*. 346–350.

[23] Fuyuan Lyu, Xing Tang, Dugang Liu, Haolun Wu, Chen Ma, Xiuqiang He, and Xue Liu. 2023. Feature representation learning for click-through rate prediction: A review and new perspectives. *arXiv preprint arXiv:2302.02241* (2023).

[24] Wentao Ouyang, Xiuwu Zhang, Lei Zhao, Jinmei Luo, Yu Zhang, Heng Zou, Zhaojie Liu, and Yanlong Du. 2020. MiNet: Mixed interest network for cross-domain click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 2669–2676.

[25] Yi Ouyang, Bin Guo, Xing Tang, Xiuqiang He, Jian Xiong, and Zhiwen Yu. 2021. Mobile app cross-domain recommendation with multi-graph neural network. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 4 (2021), 1–21.

[26] Qi Pi, Weijie Bian, Guorui Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Practice on long sequential user behavior modeling for click-through rate prediction. In *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2671–2679.

[27] Qi Pi, Guorui Zhou, Yujing Zhang, Zhe Wang, Lejian Ren, Ying Fan, Xiaoqiang Zhu, and Kun Gai. 2020. Search-based user interest modeling with lifelong sequential behavior data for click-through rate prediction. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. 2685–2692.

[28] Jiarui Qin, Weinan Zhang, Rong Su, Zhirong Liu, Weiwen Liu, Guangpeng Zhao, Hao Li, Ruiming Tang, Xiuqiang He, and Yong Yu. 2023. Learning to retrieve user behaviors for click-through rate estimation. *ACM Transactions on Information Systems* 41, 4 (2023), 1–31.

[29] Jiarui Qin, Weinan Zhang, Xin Wu, Jiarui Jin, Yuchen Fang, and Yong Yu. 2020. User behavior retrieval for click-through rate prediction. In *Proceedings of the 43rd International Conference on Research and Development in Information Retrieval*. 2347–2356.

[30] Kan Ren, Jiarui Qin, Yuchen Fang, Weinan Zhang, Lei Zheng, Weijie Bian, Guorui Zhou, Jian Xu, Yong Yu, Xiaoqiang Zhu, and Kun Gai. 2019. Lifelong sequential modeling with personalized memorization for user response prediction. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 565–574.

[31] Steffen Rendle. 2010. Factorization machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining*. 995–1000.

[32] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting clicks: Estimating the click-through rate for new ads. In *Proceedings of the 16th International Conference on World Wide Web*. 521–530.

[33] Zihua Si, Lin Guan, Zhongxiang Sun, Xiaoxue Zang, Jing Lu, Yiqun Hui, Xingchao Cao, Zeyu Yang, Yichen Zheng, Dewei Leng, Kai Zheng, Chenbin Zhang, Yanan Niu, Yang Song, and Kun Gai. 2024. TWIN V2: Scaling ultra-long user behavior sequence modeling for enhanced CTR prediction at kuaishou. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*. 4890–4897.

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*.

[35] Tianzi Zang, Yanmin Zhu, Ruohan Zhang, Chunyang Wang, Ke Wang, and Jiadi Yu. 2023. Contrastive multi-view interest learning for cross-domain sequential recommendation. *ACM Transactions on Information Systems* 42, 3 (2023), 1–30.

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

[37] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*. 5941–5948.

[38] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1059–1068.

[39] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-domain recommendation: Challenges, progress, and prospects. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*. 4721–4728.

A More Experimental Results

A.1 Efficiency Analysis

Inference latency and storage cost are critical concerns in real-world recommendation systems. Hence, in this section, we analyze the inference latency and the number of parameters of our proposed method to validate its efficiency.

Regarding inference latency, the main challenge lies in computing textual embeddings using a pre-trained language model (PLM). However, in our online deployment, we maintain a Redis table including <user, behaviors> pairs, which will be updated with item embedding in one day. We also filter the source behaviors not in our target domain candidate items during the offline stage. As a result, in the inference stage, we only retrieve the items according to the similarity metric for the user. The overall inference time is shown in Table 5, from which we can observe that the inference latency of our RAL-CDNet is comparable to that of other models.

As for the number of parameters, the PLM is used in the CD-RU module and is frozen for retrieval, thus introducing no additional trainable parameters. As we only retrieve behaviors from the source domain, the increase of parameters in the embedding layer depends on the number of retrieved behaviors compared with a single domain, which is the same in the cross-domain setting. Then, CD-AU only introduces an MLP, and Cross-Net is a typical cross-domain sequential CTR prediction network. The comparison of parameter counts is also reported in Table 5.

Table 5: Inference time and number of parameters statistics.

Dataset	Amazon	
Source → Target	Book → CD	
Models	Inference Time (s)	Number of Parameters
SIM(Soft)	0.0114	9,517,957
CoNet	0.0087	16,060,338
MiNet	0.0110	16,640,098
LCN	0.0162	16,151,435
RAL-CDNet	0.0142	16,461,074

Table 6: Sensitivity analysis of PLM similarity threshold θ, where the best results are marked in bold.

Dataset	Amazon		
Source → Target	Book → CD		
θ	AVG Length of Retrieved Sequence	AUC↑	Logloss↓
0.8	3.84	0.7273	0.3506
0.7	12.32	**0.7310**	**0.3357**
0.6	25.53	0.7298	0.3387
0.5	26.71	0.7281	0.3393

A.2 Parameter Sensitivity Analysis

The similarity threshold θ is an important hyper-parameter in our method. Here, we conduct a sensitivity analysis of θ on one dataset from Amazon, and the results are shown in Table 6.

It can be observed that the smaller the value of θ, the longer the retrieved item sequence becomes, but more noisy items are also introduced. Conversely, the larger the value of θ, the shorter the retrieved sequence is, resulting in less informative content. Thus, we use $\theta = 0.7$ as the best similarity threshold for retrieval in the CD target domain. In future work, we may consider treating θ as a learnable parameter to further explore ways to balance complexity and performance.