Sequential Recommendation via an Adaptive Cross-domain Knowledge Decomposition

Chuang Zhao*
College of Management and Economics, Tianjin University
Tianjin, China
zhaochuang@tju.edu.cn

Xinyu Li*
College of Management and Economics, Tianjin University
Tianjin, China
lxyt7642@tju.edu.cn

Ming HE†
AI Lab at Lenovo Research
Beijing, China
heming01@foxmail.com

Hongke Zhao†
College of Management and Economics, Tianjin University
Tianjin, China
hongke@tju.edu.cn

Jianping Fan
AI Lab at Lenovo Research
Beijing, China
jfan1@lenovo.com

ABSTRACT

Cross-domain recommendation, as an intelligent machine to alleviate data sparsity and cold start problems, has attracted extensive attention from scholars. Existing cross-domain recommendation frameworks usually leverage overlapping entities for knowledge transfer, the most popular of which are information aggregation and consistency maintenance. Despite decent improvements, the neglect of dynamic perspectives, the presence of confounding factors, and the disparities in domain properties inevitably constrain model performance. In view of this, this paper proposes a sequential recommendation framework via adaptive cross-domain knowledge decomposition, namely *ARISEN*, which focuses on employing adaptive causal learning to improve recommendation performance. Specifically, in order to facilitate sequence transfer, we align the user's behaviour sequences in the source domain and target domain according to the timestamps, expecting to use the abundant semantics of the former to augment the information of the latter. Regarding confounding factor removal, we introduce the causal learning technique and promote it as an adaptive representation decomposition framework on the basis of instrumental variables. For the sake of alleviating the impact of domain disparities, this paper endeavors to employ two mutually orthogonal transformation matrices for information fusion. Extensive experiments and detailed analyzes on large industrial and public data sets demonstrate that our framework can achieve substantial improvements over state-of-the-art algorithms.

*Both authors contributed equally to this research. This work was done when Chuang Zhao and Xinyu Li were doing an internship at Lenovo AI Lab.
†Corresponding authors

CIKM '23, October 21–25, 2023, Birmingham, United Kingdom

ACM ISBN 979-8-4007-0124-5/23/10…$15.00
https://doi.org/10.1145/3583780.3615058

CCS CONCEPTS

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

KEYWORDS

Casual inference, Representation decomposition, Cross-domain recommendation

ACM Reference Format:
Chuang Zhao, Xinyu Li, Ming HE, Hongke Zhao, and Jianping Fan. 2023. Sequential Recommendation via an Adaptive Cross-domain Knowledge Decomposition. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23), October 21–25, 2023, Birmingham, United Kingdom.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3583780.3615058

1 INTRODUCTION

With the data explosion of the network and mobile applications, the recommendation system, as a widely used information filtering application, has become an indispensable tool in our life [44, 54, 61, 64]. Despite its powerful capabilities in overcoming information overload and capturing user preferences, there are still certain weaknesses in addressing data sparsity [14, 23, 55] and cold-start problems [51, 60]. On account of this, a large body of research has made tremendous efforts on cross-domain recommendation using the flourishing transfer learning [24, 56, 59]. For instance, numerous companies operate not only a shopping mall platform but also an associated community platform, both serving as valuable repositories of user interests and facilitating knowledge transfer.

Existing cross-domain recommendation frameworks mainly rely on overlapping entities for knowledge transfer [8, 15], thereby improving the performance of models in single or dual domains. To be specific, the common practices of overlapping entity utilization can be mainly divided into *information aggregation* [9, 21, 27] and *consistency maintenance* [48, 49, 52, 53]. The former adopts a variety of aggregation methods, such as concat, pooling, and attention [20, 58, 59], to fuse the entity representations in the two domains, while the latter tends to attach semantic-rich regularization terms to force the embedding approximation of the same entity between different domains [34]. There are also approaches that rely on identifying domain mapping equations, which assume linear or non-linear

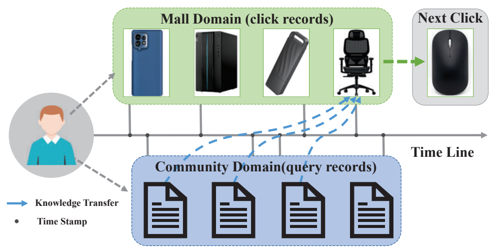


Figure 1: Examples of mall and community forum domains. Sort the click and query records of the overlap user in the mall and community alternately by timestamp. For each click record in the mall, all community query records prior to that timestamp are used as supplementary data.

relationships between different representations of domains. Our framework endeavors to promote the first paradigm.

Regardless of the decent performance, existing research on cross-domain recommendation still bears with unresolved challenges [29, 47]. First, due to the simplicity of static transfer, the vast majority of cross-domain frameworks improve on collaborative filtering or content-based recommendation. Nevertheless, these static modeling approaches can only capture general preferences of users, and there is still a lack of off-the-shelf attempts to integrate cross-domain transfer with sequential recommendation [2, 23]. Second, considering various biases in prevailing recommender systems, using coarse transfer without debiasing may introduce cumulative noise [10, 32, 63]. One potential solution to tackle this challenge is to leverage causal decoupling techniques from established single-domain recommendation methods within the cross-domain framework, as suggested in recent research studies [33, 42]. However, a drawback of this approach is that it relies on manual intervention, particularly in the selection of instrumental variables (IVs) [6, 17, 33, 46]. In particular, instrumental variables, as important elements of decoupling confounding factors and causality using two-stage least squares (2SLS), need to be manually defined in advance, whether in the traditional causal learning framework or in combination with deep learning [45]. Nonetheless, in light of recent research, it has become apparent that these IVs defined based on artificial experience may violate the conditions of effective IV during training, rendering the explicitly specified independent variable feeble or suboptimal, thus hindering the application of IV-based counterfactual prediction methods [33]. Third, there exist fundamental disparities between different domains, and the distinctive attributes of each domain may result in noise and performance degradation when attempting to directly integrate cross-domain knowledge [27, 33]. On the one hand, knowledge transfer across domains requires selectivity rather than integrity, especially in sequential recommendation scenarios; on the other hand, considering the temporal relationship in sequential recommendation, more fine-grained information filtering at the point-in-time level is required.

To address these challenges, this paper proposes a sequenti**A**l **R**ecommendation v**I**a an adaptive cros**S**-domain knowledg**E** decompositio**N**, i.e., *ARISEN*. Specifically, for cross-domain sequential recommendation, we propose a temporal point-to-point knowledge transfer approach. With reference to Figure 1, we treat the community and the mall as the source and target domains, respectively, utilizing all community interaction data preceding each purchase timestamp in the mall as supplementary knowledge to enhance the supervision signal at that specific time point. Targeted at alleviating accumulated bias derived from cofounders, we design an adaptive causal learning framework to reconstruct target domain embeddings for unbiased knowledge transfer. Particularly, we exploit search information as an instrumental variable [33, 46] and employ a two-stage least squares approach to guide the decomposition of community behavior representations. In contrast to the prior approach of directly incorporating instrumental variables [5, 35, 36], we introduce two mutual information constraints, namely *relevance* and *exclusion*, to facilitate collaborative optimization of instrumental variables. Given the notable disparities between the two domains, we integrate a latent orthogonal mapping [25, 26, 50]mechanism for better information fusion. More precisely, the adaptive nature of the orthogonal matrices enables them to determine whether to incorporate cross-domain knowledge based on the state of the two sequences at a specific time point. We conduct extensive experiments and detailed analysis on two large data sets to demonstrate the effectiveness and interpretability of our framework.

In a nutshell, this paper makes the following contributions:

- This work represents a remarkable effort on cross-domain sequential recommendation using an adaptive causal learning framework. We leverage the behavior sequences of the source domain as cross-domain knowledge to alleviate the data sparsity and cold-start dilemma of the target domain.
- With the aim of decoupling behavior representations more effectively, we design a mutual information loss function on the basis of the relevance and exclusion of instrumental variables, and jointly optimize it.
- We conduct extensive experiments and detailed analysis on two real-world industrial data sets. We have released all source code and sample data for reproducibility [1].

2 RELATED WORK

Our framework is derived from two research areas, *cross-domain recommendation* and *IV-based causal inference*. Below we will introduce their main genres and key differences from this work.

2.1 Cross-domain Recommendation

Cross-domain recommendation is applied to solve the headaches of data sparsity and cold start in traditional recommendation systems. The essential idea is to leverage the information collected in the source domain to improve the recommendation performance in the target domain [47]. Existing methods mainly include *information aggregation, consistency maintenance* and *mapping equation seeking*. For information aggregation, DTCDR [58] attempts to combine text content and ratings of overlapping users in three ways (concatenation, max-pooling, average-pooling) to achieve bidirectional transfer of user preferences. CoNet [20] shares representations of overlapping user to achieve dual-domain collaborative training. For consistency maintenance genre, CDRIB [7] designs two information bottleneck regularizers on the basis of information bottleneck principle, aiming to jointly learn cross-domain and intra-domain

[1] https://github.com/LxytIUON/ARISEN

user representations. CIT [49] draws on domain adaptive technology to maximize the overall fit of user information across domains, thereby ensuring that the knowledge extracted from the source domain fits the target domain. For methods based on mapping functions, DDTCDR [25] seeks mutual mapping equations of two domain representations to maintain domain-specific properties. PTUPCDR [61] designs a cross-domain oriented meta-learner to generate a personalized preference transfer bridge for each overlapping user, so as to realize the personalized transfer of user interests.

Our algorithm strives to improve on the first paradigm. Particularly, In contrast to prior methodologies centering on refining the process of information aggregation, the approach delineated herein is intended to acquire unbiased cross-domain knowledge via representational reconstruction, thus eradicating the prejudicial influence of confounding variables on information aggregation.

2.2 IV-based Causal Inference

Causal inference is an effective machine in scientific and commercial applications, where the IV-based paradigm is widely welcomed for its simplicity. Traditional IV-based counterfactual prediction methods mostly adopt the two-stage least squares [3], which assumes that there exists a linear relationship between random variables. There are a host of IV-based causal learning methods that extend 2SLS beyond this. GMM [16] performs parameter estimation based on the moment condition satisfied by the actual parameters of the model. DeepGMM [5] uses deep neural network to extend GMM-based methods and realize nonlinear IVs regression. KIV [35] employs kernel ridge regression to model the relationship between variables as nonlinear relationship in reproducing kernel Hilbert spaces. Deep IV [17] proposes a flexible framework to fit mixed density networks for treatment and trains an outcome prediction model with estimated conditional treatment distributions, thus combining the 2SLS with deep learning to overcome the constraints of linear relationships and dimensions. IV4Rec [33] applies the IV-based idea into cross-domain recommendation scenario and conduct representation debias by causal decoupling.

Despite decent performance, most of the aforementioned works require predefined effective IVs, which are difficult to satisfy in the real scenario due to the harshness of their conditions (such as relevance, exclusion). In view of this, this paper puts a deep insight into the conditions of instrumental variables. Moreover, novel mutual information constraints are ingeniously formulated to enable the synergistic optimization of said variables throughout the training protocol, so as to better perform causal decomposition.

3 PROPOSED METHOD

In this section, we introduce our proposed *ARISEN* framework. We first formalize the problem definition and optimization goals, then elaborate the technical details of each sub-module, and finally illustrate the model training.

3.1 Problem Formulation

Suppose there is a source domain (community) and a target domain (mall), and they share an overlapping set of users U and non-overlapping item sets $I^{\mathcal{M}}$ and I^{C}. The behavior records of users in the mall and community are $D^{\mathcal{M}}$ and D^{C} respectively,

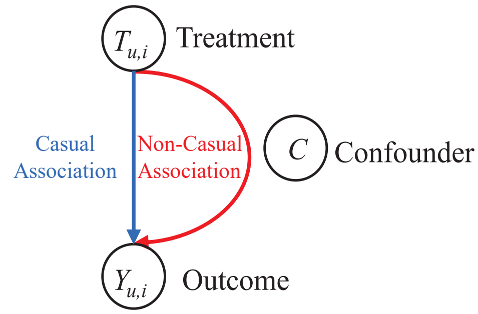


Figure 2: Recommendations from a causal perspective. The causal associations and non-causal associations are mixed between treatment and outcome.

where $(u, i, c, s) \in D^{\mathcal{M}}$ refers to the user's response to product i at time s. $c = 0$ means that u does not click i, otherwise clicks, where i refers to products and posts in the mall and community respectively. In addition, the users' searching behavior in the two domains is recorded as D^{Q}, where $(u, q, s) \in D^{Q}$ represents u's searching content q at time point s. In this paper, our goal is to reconstruct the treatment in D^{C} with the help of D^{Q} and utilize it as supplementary knowledge to enhance the recommendation performance of $D^{\mathcal{M}}$.

3.2 Recommendations from A Causal Insight

Existing sequential recommender systems are often trained by users' click histories, stemming from the belief that each user's click (u, i, c, s) unbiasedly expresses the user's preferences. Nonetheless, in reality, the user's click behavior may be disturbed by a host of confounding factors (such as position bias [22], popularity bias [62], and selection bias [28]), resulting in a decline in recommendation performance. Causal inference can be used to address this issue, and its formalization is shown in Figure 2. The user and item representations are used as joint input, also known as the treatment $T_{u,i}$, and the click behavior is under the name of the outcome $Y_{u,i}$. The confounding factor C is correlated with both $T_{u,i}$ and $Y_{u,i}$, which distorts (masks or exaggerates) the true connection between the two. In other words, the presence of C negatively affects the recommendation performance.

In cross-domain recommendation, this inferior influence may be transferred from the source domain to the target domain, and produce worse results due to the domain gap.

An intuitive solution to this dilemma is to remove the effects of confounding factors before knowledge transfer. There have been attempts of similar ideas in single-domain recommendation, which manually select an instrumental variable, and then performs 2SLS approach to decouple the representation into a causal and non-causal association part [46]. The key to the success of this scheme lies in the selection of IVs, which need to meet the following conditions [41, 45],

(1) **Relevance**: IVs Z is related to treatment T, i.e., $\mathbb{P}(T|Z) \neq \mathbb{P}(T)$.
(2) **Exclusion**: IVs Z does not directly affect outcome Y, i.e., $\mathbb{P}(Y|Z, T, C) = \mathbb{P}(Y|, T, C)$.
(3) **Unconfounded Instrument**: IVs Z should be unconfounded, i.e., $\mathbb{P}(C|Z) = \mathbb{P}(C)$.

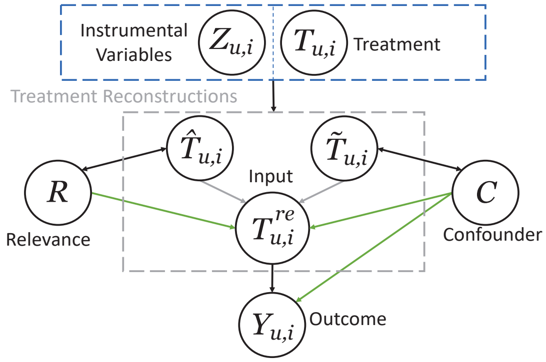


Figure 3: Treatment reconstruction. We leverage IVs to decompose treatment into causal associations and non-causal associations, and reconstruct it.

Accordingly, in our scenario, we regards user's search records in dual domains as an instrumental variable to decouple $T_{u,i}$. As depicted in Figure 3, we aggregate the user query in two domains as an IV, defined as $Z_{u,i}$, by timestamp. By regressing $T_{u,i}$ on $Z_{u,i}$ to get $\hat{T}_{u,i}$, which is not dependent on the confounders C. The relationship between $\hat{T}_{u,i}$ and the outcome $Y_{u,i}$ can be viewed as a causal association, since it is only related to the input treatment and not related to the outcome $Y_{u,i}$. We also calculated the residual $\tilde{T}_{u,i}$ of the regression, which is correlated with both the input treatment and the outcome. We regard the relationship between $\tilde{T}_{u,i}$ and outcome $Y_{u,i}$ as a non-causal association. By way of weighted combination $\hat{T}_{u,i}$ and residuals $\tilde{T}_{u,i}$, we reconstruct a more effective treatment $T^{re}_{u,i}$. In other words, we decouple the real preferences of users in the community from confounding factors and reconstruct more effective user representations.

3.3 Overview of Framework

We propose *ARISEN* to address the cross-domain transfer problem in sequential recommendation, which augments the information of the mall domain by reconstructing a more effective community domain representation and using it as cross-domain knowledge. Specifically, our framework consists of four main parts, as shown in Figure 4. In the *Variable Representation Module*, we separately define the originate treatment $T_{u,i}$ and instrumental variable $Z_{u,i}$ on the basis of causal inference and sequential recommendation. Next, we perform a representational decomposition and reconstructions of community embeddings using search records in the *Treatment Reconstruction Module*. Finally, the mall embedding and community embedding are fed into the underlying recommender model through *Orthogonal Mapping Module*. In addition, in order to alleviate the defect of manual selection of IVs, we design a training method of *IVs Collaborative Optimization* on account of its necessary condition.

3.4 Variable Representation Module

To obtain confound-free cross-domain knowledge transfer, we regard the user's interaction sequence in the source domain (community) and targeted product in the target domain (mall) as the treatment. Formally, for u,

$$T_{u,i} = \{t_j : j \in I_u^C \cup \{i\}\}, \quad i \in I_u^{\mathcal{M}}, \tag{1}$$

where t_j is the embedding vector of post j, which is usually generated by some representation learning methods (e.g., BERT [12]), and I_u^C denotes click sequence of u in the community, and $|I_u^C| = l_C$. Each t_j corresponds to a timestamp s_j^t, recording when the user u clicks on the post j.

Meanwhile, the user u also has a series of query Q_u, in which each $q_i \in Q_u$ also has a time stamp s_i^q, recording the time when the user u conducts the query q_i. As described in section 3.2, we leverage the user's search records as an instrumental variable. More precisely, for each t_j, its corresponding instrumental variable is a set of queries $(q_1, q_2, ..., q_i)$, where $s_i^q < s_j^t$. Without loss of generality, we employ cumulative concatenation to obtain instrumental variable representations. Formally,

$$q_j = (q_1 \oplus q_2 \oplus ...q_i), \quad s_i^q < s_j^t, \tag{2}$$

$$Z_{u,i} = \{z_j : j \in T_{u,i}\}, \tag{3}$$

where z_j is the embedding vector of query q_j, which is generated in the same way as treatment, i.e., through some representation learning methods (e.g., BERT).

3.5 Treatment Reconstruction Module

Once the necessary elements, i.e., $T_{u,i}$ and $Z_{u,i}$, are obtained, we perform causal decomposition using 2SLS approach to obtain reconstructed representations. In this way, we strive to obtain more efficient community representations for knowledge transfer.

3.5.1 Treatment decomposition. The essential idea of the IVs-based approach is to identify causal associations from treatment to outcome. On account of the properties of the valid IVs (i.e., IVs are not affected by confounding factors C, and only affects the outcome Y via treatments), we regress $T_{u,i}$ on $Z_{u,i}$ to get $\hat{T}_{u,i}$ (casual association) and $\tilde{T}_{u,i}$ (non-casual association). Formally,

$$\hat{T}_{u,i} = \{\hat{\mathbf{t}}_j = z_j \cdot t_j : j \in I_u^C \cup \{i\}\}, \quad i \in I_u^{\mathcal{M}}, \tag{4}$$

where $t_j \in T_{u,i}$ and $z_j \in Z_{u,i}$. The $\hat{\mathbf{t}}_j \in \hat{T}_{u,i}$ is the fitted part of the embedding t_j, which reflects the causal association between the treatment and the outcome in recommender system. Particularly, it is a closed-form solution of $\arg\min \|z_j t_j - \text{MLP}_{reg}(t_j)\|_2^2$. When $\hat{T}_{u,i}$ is obtained, the residual part of the regression $\tilde{T}_{u,i}$ can be easily obtained. Formally,

$$\tilde{T}_{u,i} = \{\tilde{\mathbf{t}}_j = \text{MLP}_{reg}(t_j) - \hat{\mathbf{t}}_j : j \in \mathcal{I}_u^C \cup \{i\}\}, \quad i \in I_u^{\mathcal{M}}. \tag{5}$$

3.5.2 Treatment reconstruction. When the treatment variables are successfully decomposed, we take the fitting vector $\hat{T}_{u,i}$, and the residual $\tilde{T}_{u,i}$ for representation reconstruction, hoping to obtain better cross-domain knowledge transfer. Fromally,

$$T^{\text{re}}_{u,i} = \{t^{\text{re}}_j = \gamma_j^1 \hat{\mathbf{t}}_j + \gamma_j^2 \tilde{\mathbf{t}}_j : j \in \mathcal{I}_u^C \cup \{i\}\}, \quad i \in I_u^{\mathcal{M}}, \tag{6}$$

where $\hat{\mathbf{t}}_j \in \hat{T}_{u,i}$ and $\tilde{\mathbf{t}}_j \in \tilde{T}_{u,i}$. We aggregate them using different weights, i.e., γ_j^1 and γ_j^2, obtained by feeding their respective representations into different transformation matrices. Formally,

$$\gamma_j^1 = W_1(\text{MLP}_{reg}(t_j) \oplus z_j), \quad \gamma_j^2 = W_2(\text{MLP}_{reg}(t_j) \oplus z_j), \tag{7}$$

where the input of two different transformation matrices, i.e., W_1 and W_2, is the concatenation of t_j and z_j. Finally, we aggregate the representations of the community domain according to the timestamps of the mall domain to facilitate knowledge transfer. This

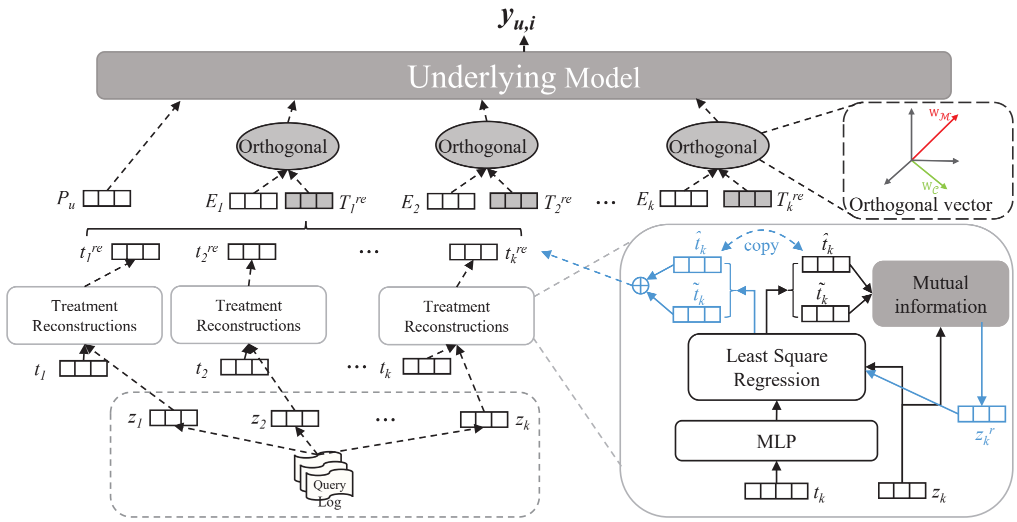


Figure 4: The overall framework of ARISEN.

approach is in line with intuition, that is, the user's behavior in the community forum serves his purchase decision in the mall [1, 30, 31]. Formally, for each $t_i^{re} \in T_{u,i}^{\text{re}}$,

$$t_i^{re} = \sum_{j=s_{i-1}^{\mathcal{M}}}^{s_i^{\mathcal{M}}} (t_j^{re}), \tag{8}$$

where $s_i^{\mathcal{M}}$ is a timestamp of i in the mall sequence. In this way, $T_{u,i}^{re}$ and $I_u^{\mathcal{M}}$ are of equal sequence length $l_{\mathcal{M}}$.

To sum up, this reconstruction is to project the non-linear representation of the embedding into the subspace spanned by the IVs, thus distinguishing the fitting part from the residual part. Furthermore, effective intervention on the fitting part and the residual part can help to mine and utilize their different mechanisms to improve the performance of the recommender system. It bears mentioning that the residual part isn't simply discarded, but rather weighted and reconstituted in light of the valid information it contained. The validity of this element will be expounded upon in section 4.3.2.

3.6 Orthogonal Mapping Module

To carry out fine-grained knowledge fusion at the time point level, we first obtain the representation of user preference in the mall domain. Formally,

$$E_{u,i} = \{e_j : j \in I_u^{\mathcal{M}}\}, \tag{9}$$

where e_j is the embedding vector of product j in mall, which is usually generated by some representation learning methods (e.g., BERT) and $|I_u^{\mathcal{M}}| = l_{\mathcal{M}}$.

In the previous section, we obtain the reconstructed community domain representation $T_{u,i}^{re}$, but the existence of domain differences makes its rough aggregation with the mall domain representation prone to noise, leading to inestimable performance degradation. In view of this, we do not directly integrate the reconstructed representation of the source domain into the target domain. Instead, we use an orthogonal mapping network to guarantee domain-specific properties. Specifically, we use two mutually orthogonal projection matrices, i.e., $W_{\mathcal{M}}$ and W_C, to transform the reconstructed source domain representation and target domain representation, expecting that the difference between the two representations is large enough while maintaining the recommendation performance. Formally,

$$< E_{u,i}, T_{u,i}^{re} > \leftrightarrow < W_{\mathcal{M}} E_{u,i}, W_C T_{u,i}^{\text{re}} >, \tag{10}$$

$$\mathcal{L}_{orth} = \sum_{i=1}^{l_{\mathcal{M}}} \frac{||e_{u,i} \cdot (t_{u,i}^{re})^{\top}||^2}{||e_{u,i}|| \cdot ||(t_{u,i}^{re})||}, \tag{11}$$

where $l_{\mathcal{M}}$ denotes length of mall sequence and $\mathcal{L}_{orth}$ is used to guarantee the orthogonalization of the two domain representations. Finally, the embeddings of the two domains are mixed into the underlying model. Formally,

$$\hat{y}_{u,i} = UnderlyingModel(P^u \oplus E_{u,i}^{\mathcal{M}} \oplus T_{u,i}^{C}), \tag{12}$$

where P^u represents user attributes. Without loss of generality, we choose DIN [57] as the underlying model in our experiments.

In a nutshell, this method has several advantages. On the one hand, the orthogonal mapping retains the special characteristics of the domain; on the other hand, compared with random initialization, it serves as a regularization term that effectively mitigates model overfitting [11, 25, 26, 50].

3.7 IVs Collaborative Optimization Module

Considering that pre-defined manual IVs may become brittle or suboptimal due to condition violations during training process, we propose a collaborative optimization method for IVs. Specifically, we exploit mutual information constraints to force IVs to satisfy *relevance*, *exclusion* and *unconfounded instrument* conditions, thereby guaranteeing the accuracy of downstream counterfactual prediction tasks. Benefiting from the previous work that has proved the exogenous nature of user search records as IVs [33], and the learned representations always satisfy the unconfounded instrument condition [5, 17, 35], we only need to make the IVs representation satisfy the relevance condition with the treatment and the exclusion condition with the outcome during the whole training.

Relevance condition. To encourage the representation of IVs $Z_{u,i}$ relevant to treatment $T_{u,i}$, we maximize the mutual information between the IVs $Z_{u,i}$ and treatment $T_{u,i}$. Formally,

$$\mathcal{L}_{ZR} = -\frac{1}{l_C l_Q}\sum_{j=1}^{l_C}\sum_{q=1}^{l_Q}(\log(MI(t_j \mid z_j)) - \log(MI(t_q \mid z_j))), \quad (13)$$

where $MI(\cdot)$ stands for mutual information estimator MINE [4], $\log(MI(t_j \mid z_j))$ represents the conditional log-likelihood of the pairs of positive samples, and $\log(MI(t_q \mid z_j))$ represents the conditional log-likelihood of the pairs of negative samples. l_Q is the size of IVs sets, equal to l_C.

Exclusion condition. In this case, the IVs $Z_{u,i}$ should be independent of the outcome $Y_{u,i}$ caused by treatment $T_{u,i}$. Instead of maximizing mutual information in learning relevance, we make positive $(y_j \mid z_j)$ and negative $(y_q \mid z_j)_{(q\neq j)}$ sample pairs have close log-likelihood expectations so that IVs and outcome Y are conditionally independent. Formally,

$$\mathcal{L}_{ZE} = \frac{1}{l_C l_Q}\sum_{j=1}^{l_C}\sum_{q=1}^{l_Q}(\log(MI(y_j \mid z_j) - \log(MI(y_q \mid z_j))). \quad (14)$$

Capitalizing on the aforementioned relevance and exclusion conditional loss functions, this approach accomplish the enhancement of similarity between the IVs and treatments concurrent with the attenuation of similarity between the former and outcomes over the course of training. This not only alleviates the defects of the traditional method due to manual definition of IVs, but also avoids the possible violation of instrumental variable conditions in the training process.

3.8 Model Training

Supervision loss. We adopt cross-entropy to guarantee the performance of the model on the target domain. Formally,

$$\mathcal{L}_{crp} = -\frac{1}{|\mathcal{D}^{\mathcal{M}}|}\sum_{(u,i,c)\in\mathcal{D}^{\mathcal{M}}} c\cdot\log\hat{y}_{u,i} + (1-c)\cdot\log(1-\hat{y}_{u,i}), \quad (15)$$

where $\hat{y}_{u,i}$ is the predicted matching score of (u, i).

Total loss. *ARISEN* is optimized in an end-to-end manner. The total loss consists of four parts: supervised loss, mutual information loss, orthogonalization constraints, and model regularization. Formally,

$$\mathcal{L}_{total} = \mathcal{L}_{crp} + \lambda_1(\mathcal{L}_{ZR} + \mathcal{L}_{ZE}) + \lambda_2\mathcal{L}_{orth} + \lambda_3\|\Theta\|^2, \quad (16)$$

where $\|\Theta\|^2$ is the regularization term to avoid overfitting. λ_1, λ_2, λ_3 are the trade-offs among four parts. *Please note that the tuning of λ_1 serves to balance the training dynamics of the IVs optimization and primary recommendation tasks.*

The parameters of *ARISEN* are derived from MLP_{reg}, W, MINE, and underlying model. All these trainable parameters are denoted as Θ, and are trained using gradient descent. Formally,

$$\Theta^{new} = \Theta^{old} - \eta\frac{\partial\mathcal{L}_{\text{total}}}{\partial\Theta^{old}}, \quad (17)$$

where η denotes learning rate.

For the complete algorithm framework, see Algorithm 1.

3.9 Plug-in Application

Prevalent sequential recommendation models [39, 40] adopt similar architecture and training methods, which utilize fixed-length user history interaction and contextual information to capture user preferences, and then employ cross-entropy or BPR loss for optimization. Our proposed framework is a plug-in framework that can be implemented on existing recommender systems by adding treatment reconstruction and IVs collaborative optimization to achieve better cross-domain knowledge transfer. Specifically, assuming that there are two domains $\mathcal{S}$ and $\mathcal{T}$, we can reconstruct the treatment of the source domain according to the constructed causal graph mentioned in section 3.2, and integrate them with the information of the target domain to obtain a semantic-rich representation X_u^{re}, thereby improving the recommendation performance. Formally,

$$X_u^{re} = Agg((Agg(E_{u,i}^{\mathcal{T}}, T_{u,i}^{\mathcal{S}}), P^u)), \quad (18)$$

where P^u refers to contextual information such as user tags and user profiles, and Agg can be any aggravated function, such as concatenation, mean pooling or attention mechanism. Finally, the matching score can be predicted based on the learned representations of user and items. Formally,

$$\hat{y}_{u,i} = f_{pred}(X_u^{re}, y_i), \quad (19)$$

where $f_{pred}(\cdot)$ can be any model that predicts the matching score, such as dot product or MLP layer. We will give specific experiments to prove this point in the experimental section.

The primary role of the ARISEN is to introduce causal associations in the resource domain so as to strengthen the user representations in the target domain. Therefore, in addition to the sequence recommendation task, it can be aggregated with other recommendation tasks using user representations in the form of modifying interfaces, such as session-based recommendations, collaborative filtering, and so on.

Algorithm 1 The Algorithm of *ARISEN*

Input: Mall click sequence $I_u^{\mathcal{M}}$, Community click sequence $I_u^{\mathcal{C}}$, Dual domain query sequence Q_u; Target product y_i.
Output: Model parameters Θ;

```
Random initialize model parameters Θ;
Data preprocessing ;
while not converged do
    Sample a batch of training data;
    Constructing treatment T_{u,i} and IVs Z_{u,i};
    Obtain causal association T̂_{u,i} as equation 4;
    Latent orthogonal mapping as equation 11;
    Feed into underlying model as equation 12;
    IVs co-optimization as equation 13 and 14;
    Caculate total loss as equation 16;
    Update parameters Θ as equation 17.
end while
return Parameters Θ
```

4 EXPERIMENTS

In this section, we conduct experiments on a large real industrial data set to evaluate the performance of our proposed model, aiming to answer the following research questions (RQs):

- RQ1: How does *ARISEN* perform compared to existing baselines?
- RQ2: How robust and interpretable is *ARISEN*?
- RQ3: How do key hyperparameters affect *ARISEN*?

4.1 Experimental Settings

4.1.1 Data set. In our experiments, we utilize a real-world large-scale industrial data set as well as a publicly available Amazon [2] data set to evaluate the performance of the model. Table 1 shows necessary statistics for these data sets.

The large-scale industrial data set is collected from behavior logs of overlapping users in *shopping malls* (source domain) and *community forums* (target domain). The data time frame for both platforms is from January to November 2022. **For the Amazon data set**, we construct cross-domain tasks using two subcategories, *Book* (source domain) and *Movies and TV* (target domain). As for there is no record of user queries, we treat user comment data as queries. In order to ensure data quality and facilitate model training, we filter out users whose click sequence length is less than 5 according to [33], and cut all behavior sequences to a length of 50. In particular, for sequence whose length is less than 50, we use 0 for padding, and for sequences whose length is greater than 50, we cut them into multiple sequences. We leverage BERT [12] to process user/product profiles and post content separately to obtain initial embeddings.

Table 1: Statistics of the industrial and amazon data set.

data set	Industrial		Amazon	
	mall	community	movie	book
users	4,391	4,391	35,146	35,146
Items	3,090	1,143	77,522	514,808
Interactions	206,335	15,046	621,008	1,167,757
Sparsity	98.48%	99.70%	99.98%	99.99%
Queries	25,808		534,688	

Our data set splitting adopts the leave-one-out method widely used in sequential recommendation [40], taking the last item of the sequence as the test set, the penultimate item as the validation set, and the rest as the training data. For training, we equip each ground truth for training with 7 negative samples to enhance the discriminative ability. For performance evaluation, we mix the ground truth of the test set into randomly selected 99 items that have not been clicked, and then rank them according to the matching scores of users and these items.

4.1.2 Baselines. We compare *ARISEN* with the state-of-the-art recommendation models. More specifically, these baselines can be categorized into *single-domain recommendation models* (NMF, DIN, IV4Rec) and *cross-domain recommendation models* (CDRIB, CoNet, DTCDR, DDTCDR). Note that for the single-domain recommendation baseline, only the target domain data is used for training and performance reporting, while for the cross-domain recommendation baseline, we use the data of both domains for training and then report the model performance on the target domain.

- NMF [18]: NMF focuses on both linear and nonlinear relationships between users and items.
- DIN [57]: DIN employs an attention mechanism to mine users' interest from their historical interaction.
- IV4Rec [33]: IV4Rec leverages search logs to reconstruct user representations in a causal manner.
- MVDNN [13]: MVDNN shares user representation extractors across multiple domains for interest transfer.
- CDRIB [7]: CDRIB designs two regularizers, achieving debiasing and capture cross-domain user dependencies.
- CoNet [20]: CoNet establishes cross-merging between the twin-tower models of both domains for knowledge transfer.
- DTCDR [58]: DTCDR designs an adaptive embedding sharing strategy based on multi-task learning (MTL).
- DDTCDR [25]: DDTCDR utilizes the two-way mapping equation to realize the interest transfer of cold-start users.

For fairness, we carefully fine-tuned all baselines to achieve their best performance.

4.1.3 Evaluation metrics and parameter settings. The data preprocessing and train-test splitting strategy have been described in detail in section 4.1.1. The evaluation metrics used in this study include Area Under Curve (AUC), Hit Ratio (HR@N), Normalized Discounted Cumulative Gain (NDCG@N), and Mean Reciprocal Rank (MRR) [47]. HR@N is employed to assess the recall ability, while AUC and NDCG@N are utilized to measure the ranking ability. In particular, we use MRR to measure the ranking ability of the entire data set, i.e., $N = |I^{\mathcal{M}}|$. For all metrics, higher is better.

Experiments were performed on an Ubuntu 18.04 server with an Intel(R) Xeon(R) Gold 5118 CPU (12 cores, 2.30GHz) and a single NVIDIA Tesla V100 GPU. We implement the model using PyTorch and optimize the parameters using the Adam optimizer with a learning rate of 5e-4 and weight decay of 5e-4. The training is performed for 100 epochs with a batch size of 1024. Hyperparameter search is conducted for essential parameters to optimize their values. Specifically, the embedding size rate is 128, derived from $[32, 64, 128, 256]$. The mutual information weight is 0.1, obtained from $[0.1, 0.5, 1, 5, 10]$. The sequence length is 50, which is searched from $[10, 30, 50, 70, 100]$. See section 4.4 for their tuning process.

4.2 Comparison with Baselines (RQ1)

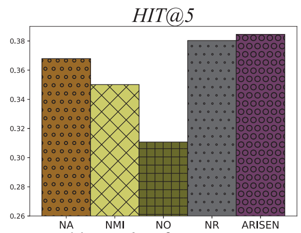
To answer RQ1 and verify the validity of our model, we compare the performance of our model to the advanced baselines. Table 2 shows the results of all algorithms on the industrial data set and the Amazon data set, respectively. Clearly, *ARISEN* outperforms all baselines on any metrics, demonstrating its effectiveness and progressiveness. Specifically, compared with the traditional single-domain recommendation algorithms NMF and DIN, cross-domain recommendation algorithms such as *ARISEN*, CDRIB and DTCDR have greatly improved recall and ranking, showing the importance of cross-domain knowledge integration. Compared with other cross-domain recommendation algorithms, *ARISEN* is 2.179%, 4.005% better than the best cross-domain recommendation model in HIT@5 and NDCG@5, and 2.272% better than the best cross-domain recommendation model in NDCG@10. In MRR, a metric that uses the entire item set test, *ARISEN* still has a large advantage, with respective improvements of 4.186% and 1.959% in two data sets compared to the best baseline. Particularly, MVDNN and DDTCDR perform unwell, even lower than single-domain recommendation algorithms on some metrics, which proves that coarse cross-domain knowledge sharing and ingenuous mapping equations may have negative impacts. IV4Rec adopts causal reconstruction similar to

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

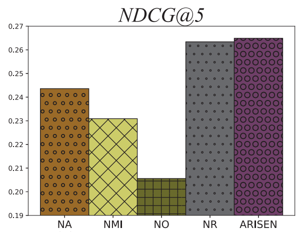
Table 2: Performance comparison with baselines.

	Industrial							Amazon						
ALGORITHM	AUC	HIT\NDCG@1	HIT@5	HIT@10	NDCG@5	NDCG@10	MRR	AUC	HIT\NDCG@1	HIT@5	HIT@10	NDCG@5	NDCG@10	MRR
NMF	0.7695	0.0954	0.3005	0.4330	0.1981	0.2409	0.2038	0.5901	0.0361	0.1125	0.1851	0.0741	0.0974	0.0946
DIN	0.8145	0.1137	0.3679	0.5114	0.2436	0.2900	0.2415	0.7543	0.1730	0.3896	0.4914	0.2861	0.3190	0.2817
IV4Rec	0.8203	0.1262	0.3763†	0.5323†	0.2533	0.3037†	0.2526	0.7579	0.1732	0.3836	0.4876	0.2828	0.3131	0.2786
MVDNN	0.7597	0.0811	0.2754	0.4235	0.1765	0.2241	0.1843	0.7274	0.1622	0.3710	0.4770	0.2706	0.3049	0.2675
CoNet	0.8206	0.1187	0.3623	0.5234	0.2438	0.2957	0.2466	0.7606	0.1531	0.3814	0.5083	0.2707	0.3116	0.2682
DTCDR	0.8208	0.1173	0.3663	0.5253	0.2431	0.2943	0.2440	0.7657	0.1756†	0.4099†	0.5297†	0.2966†	0.3344†	0.2905†
DDTCDR	0.7959	0.1301†	0.3708	0.5029	0.2547†	0.2974	0.2532†	0.6228	0.0718	0.2027	0.2778	0.1388	0.1631	0.1478
CDRIB	0.8227†	0.1147	0.3634	0.5285	0.2395	0.2926	0.2417	0.7716†	0.1367	0.3660	0.4980	0.2547	0.2973	0.2538
ARISEN	**0.8236**	**0.1371**	**0.3845**	**0.5341**	**0.2649**	**0.3106**	**0.2638**	**0.7780**	**0.1824**	**0.4194**	**0.5321**	**0.3052**	**0.3378**	**0.2962**
Improvements	0.109%	5.380%	2.179%	0.338%	4.005%	2.272%	4.186%	0.822%	3.853%	2.330%	0.466%	2.923%	1.018%	1.959%

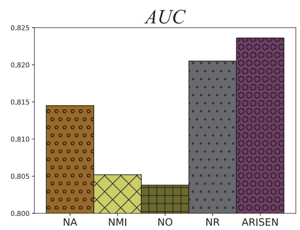
Improvements means the relative improvement of ARISON compared to the optimal baseline †.



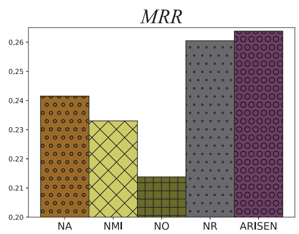
(a) Results of HIT@5.



(b) Results of NDCG@5.

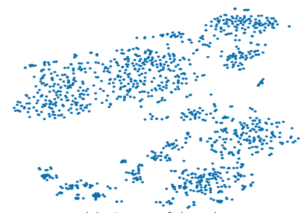


(c) Results of AUC.

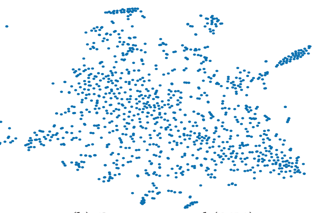


(d) Results of MRR

Figure 5: Ablation results.



(a) Original (DIN).



(b) Reconstrcuted (DIN).

(c) Original (SSE-PT).

(d) Reconstrcuted (SSE-PT).

Figure 6: T-SNE visualization of the treatment variable.

ours, but its single-domain limitation and non-optimization of instrumental variables make its performance weaker than *ARISEN*. It is worth noting that *ARISEN* has a greater improvement in more difficult scenarios, namely Hit@1, NDCG@1, which further proves its advanced nature. To sum up, cross-domain algorithms are stronger than single-domain algorithms, and fine-grained knowledge integration and IVs-cooptimization are better than plain information aggregation. *ARISEN* takes its essence and discard its dross.

4.3 Robust Tests (RQ2)

To answer RQ2, without loss of generality, we perform extensive robustness tests on the industrial data set, such as plug-in application, ablation experiments, and visualizations.

4.3.1 Plug-in application. To demonstrate the generality and broad prospects of our innovation, we improve on the classic sequential recommendation model, i.e., GRU4Rec [19] , CASER [37], SSE-PT [43]. These underlying models are proven to achieve good performance on sequential recommendation.

As shown in Table 3, adding our innovations to all underlying models can promote their metrics. Specifically, on industrial data set, *ARISEN* enhances GRU4Rec, improving recall and ranking performance by up to 6.415%, 4.970%, respectively. Meanwhile, our innovations have greatly improved in CASER and SSE-PT, and have increased by more than 10% in many metrics. We also conduct similar tests on the Amazon data set, and the plugin demonstrate significant improvements for the original algorithm. These results fully demonstrate the effectiveness and pluggability of our innovations.

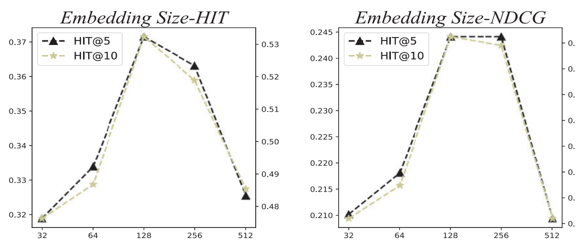
4.3.2 Ablation study. To test the effectiveness of each sub-module, we perform module elimination and keep other parts unchanged.

- ARISEN-NA: This variant eliminates all cross-domain-related innovations and utilizes only the underlying models.
- ARISEN-NMI: This variant eliminates co-optimization of IVs.
- ARISEN-NO: It directly fuses representations without considering domain disparties.
- ARISEN-NR: It only uses causal representations $\hat{T}_{u,i}$ for fusion.

As shown in Figure 5, the performance of our model is higher than that of the variants in any metrics, which fully proves the indispensability of each sub-module. To be specific, ARISEN-NA suffers from unsatisfactory result due to lack of cross-domain knowledge and data sparsity. ARISEN-NO ignores domain differences and directly performs information fusion, which results in noise interference and performance degradation. ARISEN-NMI has inferior performance because it discards the co-optimization of instrumental variables. This approach not only fails to alleviate the inaccuracy of manual selection of instrumental variables, but also fails to properly decouple causality as a consequence of the inability to maintain relevance and exogenous conditions during the optimization process. ARISEN-NR is slightly weaker than our results, because confounding factors and outcome are correlated [33], thus cannot be discarded directly. In a nutshell, our model comprehensively considers the rationality of instrumental variables and domain differences, and improves performance through targeted model design.

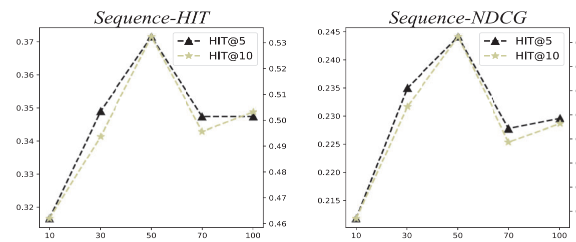
Table 3: Experimental results of plug-in applications.

	Industrial							Amazon						
ALGORITHM	AUC	HIT\NDCG@1	HIT@5	HIT@10	NDCG@5	NDCG@10	MRR	AUC	HIT\NDCG@1	HIT@5	HIT@10	NDCG@5	NDCG@10	MRR
GRU4Rec	0.7826	0.0966	0.3046	0.4474	0.2014	0.2475	0.2075	0.7628	0.1801	0.4190	0.5185	0.3046	0.3370	0.2952
ARISEN-GRU4Rec	**0.8037**	**0.1002**	**0.3112**	**0.4761**	**0.2067**	**0.2598**	**0.2156**	**0.7776**	**0.1850**	**0.4244**	**0.5378**	**0.3070**	**0.3436**	**0.2985**
Improvements	2.696%	3.727%	2.167%	6.415%	2.632%	4.970%	3.904%	1.944%	2.721%	1.294%	3.717%	0.788%	1.970%	1.105%
CASER	0.7418	0.0606	0.2162	0.3419	0.1389	0.1794	0.1551	0.7683	0.1608	0.3970	0.5094	0.2835	0.3199	0.2774
ARISEN-CASER	**0.7533**	**0.0697**	**0.2376**	**0.3850**	**0.1535**	**0.2009**	**0.1684**	**0.7720**	**0.1792**	**0.4054**	**0.5130**	**0.2952**	**0.3300**	**0.2894**
Improvements	1.550%	15.017%	9.898%	12.606%	10.511%	11.984%	8.575%	0.481%	11.431%	2.111%	0.693%	4.136%	3.167%	4.328%
SSE-PT	0.7939	0.1039	0.3189	0.4656	0.2127	0.2602	0.2184	0.7716	0.1813	0.4228	0.5265	0.3070	0.3404	0.2978
ARISEN-SSE-PT	**0.8249**	**0.1221**	**0.3793**	**0.5292**	**0.2512**	**0.2997**	**0.2492**	**0.7755**	**0.1840**	**0.4264**	**0.5323**	**0.3101**	**0.3441**	**0.3008**
Improvements	3.905%	17.517%	18.940%	13.660%	18.101%	15.181%	14.103%	0.502%	1.472%	0.841%	1.108%	0.987%	1.076%	1.014%



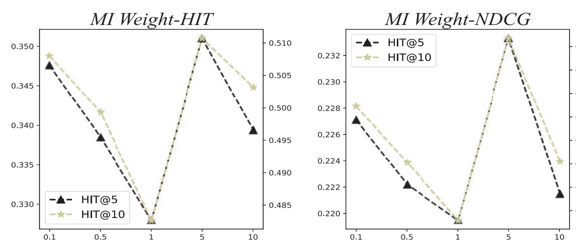
(a) Results of HIT. (b) Results of NDCG.

Figure 7: Hyperparameter test - embedding size



(a) Results of HIT. (b) Results of NDCG.

Figure 8: Hyperparameter test - sequence length



(a) Results of HIT. (b) Results of NDCG.

Figure 9: Hyperparameter test - mutual information weight

4.3.3 Treatment variable visualization. To further validate the effectiveness and understandability of our causal decoupling module, we also show visualization of the treatment variable before and after reconstruction using T-SNE [38].

Figure 6(a) and Figure 6(b) are visualizations of the treatment variable before and after causal decoupling in ARISEN, while Figure 6(c) and Figure 6(d) are for ARISEN-SSE-PT. By comparison, there is evidence that the reconstructed embeddings are more evenly distributed than the original ones, which is attributed to the fact that *ARISEN* generally removes the effects of confounding factors.

4.4 Hyper-testing (RQ3)

To answer RQ3, we investigate the impact of key hyperparameter settings on model performance on industrial data set. Specifically, we present the impact of embedding size k, sequence length $l_{\mathcal{M}}$, and mutual information weight λ_1 on *ARISEN*, respectively.

Embedding size k. Embedding size represents model capacity and complexity. Small k may cause underfitting of the model, and vice versa, overfitting [58]. As shown in Figure 7, our model achieves the best performance when $k = 128$.

Sequence length $l_{\mathcal{M}}$. The sequence length refers to the history data used for training. Longer sequences imply more complete patterns and interest preferences, but may bring noise. The evidence in Figure 8 shows that the model works best when $l = 50$.

Mutual information weight λ_1. Mutual information weight is a trade-off between IVs co-optimizations and recommendation performance. A larger weight tends to satisfy the conditions of instrumental variables, and vice versa, more attention is paid to the fitting of recommendation results. Figure 9 shows that our proposed model is better when $\lambda_1 = 5$.

5 CONCLUSION

In this paper, on the basis of causal learning techniques, we propose a sequential recommendation framework via adaptive cross-domain knowledge decomposition (i.e., *ARISEN*) to boost recommendation performance. *ARISEN* employs the user's search records in the two domains as instrumental variables to decompose their interaction behavior in the community domain into purchase-related and purchase-independent parts. Meanwhile, in order to overcome the arbitrariness and labor cost of instrumental variable selection, *ARISEN* incorporates an adaptive learning paradigm that utilizes mutual information maximization for instrumental variable collaborative optimization. Two orthogonal matrices are applied to information fusion to alleviate possible noise and performance degradation caused by domain characteristics. Extensive comparative experiments and robustness tests with state-of-the-art baselines demonstrate the effectiveness and interpretability of our model. Despite this, our model still has some limitations that are left for future work. First, our innovations are pluggable applications and should be tested on more underlying models. Second, we should test the performance of our model on more large-scale cross-domain tasks, especially in scenarios with large domain differences.

ACKNOWLEDGMENTS

This study was funded by the supports of National Natural Science Foundation of China (72101176).

REFERENCES

[1] Ali Abdallah Alalwan. 2018. Investigating the impact of social media advertising features on customer purchase intention. *International Journal of Information Management* 42 (2018), 65–77.

[2] Chainarong Amornbunchornvej, Elena Zheleva, and Tanya Berger-Wolf. 2021. Variable-lag granger causality and transfer entropy for time series analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 4 (2021), 1–30.

[3] Joshua D Angrist and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

[4] Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. 2018. Mutual information neural estimation. In *International conference on machine learning*. PMLR, 531–540.

[5] Andrew Bennett, Nathan Kallus, and Tobias Schnabel. 2019. Deep generalized method of moments for instrumental variable analysis. *Advances in neural information processing systems* 32 (2019).

[6] Mehmet Caner and Bruce E Hansen. 2004. Instrumental variable estimation of a threshold model. *Econometric Theory* 20, 5 (2004), 813–843.

[7] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-domain recommendation to cold-start users via variational information bottleneck. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2209–2223.

[8] Chaochao Chen, Huiwen Wu, Jiajie Su, Lingjuan Lyu, Xiaolin Zheng, and Li Wang. 2022. Differential Private Knowledge Transfer for Privacy-Preserving Cross-Domain Recommendation. In *Proceedings of the ACM Web Conference 2022*. 1455–1465.

[9] Chong Chen, Min Zhang, Chenyang Wang, Weizhi Ma, Minming Li, Yiqun Liu, and Shaoping Ma. 2019. An efficient adaptive transfer neural network for social-aware recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 225–234.

[10] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2020. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems* (2020).

[11] Zhihong Chen, Jiawei Wu, Chenliang Li, Jingxu Chen, Rong Xiao, and Binqiang Zhao. 2022. Co-training Disentangled Domain Adaptation Network for Leveraging Popularity Bias in Recommenders. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 60–69.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv* abs/1810.04805 (2019).

[13] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th international conference on world wide web*. 278–288.

[14] Wenjing Fu, Zhaohui Peng, Senzhang Wang, Yang Xu, and Jin Li. 2019. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 94–101.

[15] 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. *arXiv preprint arXiv:2105.03300* (2021).

[16] Lars Peter Hansen. 1982. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society* (1982), 1029–1054.

[17] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. 2017. Deep IV: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*. PMLR, 1414–1423.

[18] 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.

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

[20] 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.

[21] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. MTNet: a neural approach for cross-domain recommendation with unstructured text. *KDD Deep Learning Day* (2018), 1–10.

[22] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the tenth ACM international conference on web search and data mining*. 781–789.

[23] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-supervised learning for cross-domain recommendation to cold-start users. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1563–1572.

[24] Adit Krishnan, Mahashweta Das, Mangesh Bendre, Hao Yang, and Hari Sundaram. 2020. Transfer learning via contextual invariants for one-to-many cross-domain recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1081–1090.

[25] Pan Li and Alexander Tuzhilin. 2020. Ddtcdr: Deep dual transfer cross domain recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 331–339.

[26] Pan Li and Alexander Tuzhilin. 2021. Dual Metric Learning for Effective and Efficient Cross-Domain Recommendations. *IEEE Transactions on Knowledge and Data Engineering* 35 (2021), 321–334. https://api.semanticscholar.org/CorpusID:233296239

[27] Siqing Li, Liuyi Yao, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Tonglei Guo, Bolin Ding, and Ji-Rong Wen. 2021. Debiasing Learning based Cross-domain Recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 3190–3199.

[28] Zohreh Ovaisi, Ragib Ahsan, Yifan Zhang, Kathryn Vasilaky, and Elena Zheleva. 2020. Correcting for selection bias in learning-to-rank systems. In *Proceedings of The Web Conference 2020*. 1863–1873.

[29] Yu Peng. 2022. A Survey on Modern Recommendation System based on Big Data. *ArXiv* abs/2206.02631 (2022).

[30] Gerard Prendergast, David Ko, and V Yuen Siu Yin. 2010. Online word of mouth and consumer purchase intentions. *International journal of advertising* 29, 5 (2010), 687–708.

[31] Muhammad Usman Riaz, Luo Xiao Guang, Maria Zafar, Fakhar Shahzad, Muhammad Shahbaz, and Majid Lateef. 2021. Consumers' purchase intention and decision-making process through social networking sites: a social commerce construct. *Behaviour & Information Technology* 40, 1 (2021), 99–115.

[32] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*. PMLR, 1670–1679.

[33] Zihua Si, Xueran Han, Xiao Zhang, Jun Xu, Yue Yin, Yang Song, and Ji-Rong Wen. 2022. A Model-Agnostic Causal Learning Framework for Recommendation using Search Data. In *Proceedings of the ACM Web Conference 2022*. 224–233.

[34] 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.

[35] Rahul Singh, Maneesh Sahani, and Arthur Gretton. 2019. Kernel Instrumental Variable Regression. In *Neural Information Processing Systems*.

[36] Gokul Swamy, Sanjiban Choudhury, J. Andrew Bagnell, and Zhiwei Steven Wu. 2022. Causal Imitation Learning under Temporally Correlated Noise. In *International Conference on Machine Learning*.

[37] 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.

[38] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[39] Shoujin Wang, Longbing Cao, Yan Wang, Quan Z Sheng, Mehmet A Orgun, and Defu Lian. 2021. A survey on session-based recommender systems. *ACM Computing Surveys (CSUR)* 54, 7 (2021), 1–38.

[40] Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z Sheng, and Mehmet Orgun. 2019. Sequential recommender systems: challenges, progress and prospects. *arXiv preprint arXiv:2001.04830* (2019).

[41] Yixin Wang, Dawen Liang, Laurent Charlin, and David M Blei. 2020. Causal inference for recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*. 426–431.

[42] Tianxin Wei, Fuli Feng, Jiawei Chen, Ziwei Wu, Jinfeng Yi, and Xiangnan He. 2021. Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 1791–1800.

[43] Liwei Wu, Shuqing Li, Cho-Jui Hsieh, and James Sharpnack. 2020. SSE-PT: Sequential recommendation via personalized transformer. In *Fourteenth ACM Conference on Recommender Systems*. 328–337.

[44] Ruobing Xie, Qi Liu, Liangdong Wang, Shukai Liu, Bo Zhang, and Leyu Lin. 2022. Contrastive cross-domain recommendation in matching. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4226–4236.

[45] Liuyi Yao, Zhixuan Chu, Sheng Li, Yaliang Li, Jing Gao, and Aidong Zhang. 2021. A survey on causal inference. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 5 (2021), 1–46.

[46] Junkun Yuan, Anpeng Wu, Kun Kuang, Bo Li, Runze Wu, Fei Wu, and Lanfen Lin. 2022. Auto IV: Counterfactual Prediction via Automatic Instrumental Variable Decomposition. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16, 4 (2022), 1–20.

[47] 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.

[48] Qian Zhang, Jie Lu, Dianshuang Wu, and Guangquan Zhang. 2018. A cross-domain recommender system with kernel-induced knowledge transfer for overlapping entities. *IEEE transactions on neural networks and learning systems* 30, 7

(2018), 1998–2012.

[49] Qian Zhang, Dianshuang Wu, Jie Lu, Feng Liu, and Guangquan Zhang. 2017. A cross-domain recommender system with consistent information transfer. *Decision Support Systems* 104 (2017), 49–63.

[50] Xiaokun Zhang, Hongfei Lin, Bo Xu, Chenliang Li, Yuan Lin, Haifeng Liu, and Fenglong Ma. 2022. Dynamic intent-aware iterative denoising network for session-based recommendation. *Information Processing & Management* 59, 3 (2022), 102936.

[51] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-domain recommendation for cold-start users via aspect transfer network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 229–238.

[52] Chuang Zhao, Hongke Zhao, Ming HE, Jian Zhang, and Jianping Fan. 2023. Cross-Domain Recommendation via User Interest Alignment. In *Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) *(WWW '23)*. Association for Computing Machinery, New York, NY, USA, 887–896. https://doi.org/10.1145/3543507.3583263

[53] Chuang Zhao, Hongke Zhao, Runze Wu, Qilin Deng, Yu Ding, Jianrong Tao, and Changjie Fan. 2022. Multi-Dimensional Prediction of Guild Health in Online Games: A Stability-Aware Multi-Task Learning Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 4371–4378.

[54] Hongke Zhao, Xinpeng Wu, Chuang Zhao, Lei Zhang, Haiping Ma, and Fan Cheng. 2021. CoEA: A cooperative–competitive evolutionary algorithm for bidirectional recommendations. *IEEE Transactions on Evolutionary Computation* 26, 1 (2021), 28–42.

[55] Hongke Zhao, Chuang Zhao, Xi Zhang, Nanlin Liu, Hengshu Zhu, Qi Liu, and Hui Xiong. 2023. An Ensemble Learning Approach with Gradient Resampling for Class-Imbalance Problems. *INFORMS Journal on Computing* (2023).

[56] Lili Zhao, Sinno Jialin Pan, and Qiang Yang. 2017. A unified framework of active transfer learning for cross-system recommendation. *Artificial Intelligence* 245 (2017), 38–55.

[57] 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 international conference on knowledge discovery & data mining*. 1059–1068.

[58] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. 2019. Dtcdr: A framework for dual-target cross-domain recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1533–1542.

[59] Feng Zhu, Yan Wang, Chaochao Chen, Guanfeng Liu, and Xiaolin Zheng. 2020. A Graphical and Attentional Framework for Dual-Target Cross-Domain Recommendation.. In *IJCAI*. 3001–3008.

[60] 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.

[61] 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.

[62] Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee. 2021. Popularity bias in dynamic recommendation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. 2439–2449.

[63] Ziwei Zhu, Yun He, Xing Zhao, and James Caverlee. 2022. Evolution of Popularity Bias: Empirical Study and Debiasing. *ArXiv* abs/2207.03372 (2022).

[64] Ziwei Zhu, Jingu Kim, Trung Nguyen, Aish Fenton, and James Caverlee. 2021. Fairness among new items in cold start recommender systems. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 767–776.