A VAE-based User Preference Learning and Transfer Framework for Cross-domain Recommendation

Tong Zhang, Chen Chen, Dan Wang, *Member, IEEE,* Jie Guo, *Member, IEEE,* Bin Song, *Senior Member, IEEE*

Abstract—The core idea of cross-domain recommendation is to alleviate the problem of data scarcity. Previous methods have made brilliant successes. However, many of them mainly focus on learning an ideal mapping function across-domains, ignoring the user preferences within a specific domain, which leads to suboptimal results. In this paper, we propose a Cross-Domain Recommendation Variational AutoEncoder framework (CDRVAE), a novel extension of a variational autoencoder on cross-domain recommendations for user behaviour distribution modeling. It applies a new hybrid architecture of VAE as the backbone and simultaneously constructs two information flows, within-domain and cross-domain modeling. For the former, an asymmetric codec structure is designed to reconstruct preference distribution from domain-specific latent factors. To relieve the posterior collapse dilemma, a combined prior is employed to increase the distribution complexity. The equivalent transition by a transformation matrix and the unobserved interaction generation by cross-domain reconstruction contribute to the latter. We combine all the above components for the more accurate and reliable user features. Extensive experiments are conducted on three public benchmark datasets to validate the effectiveness of the proposed CDRVAE. Experimental results demonstrate that CDRVAE is consistently superior to other state-of-the-art alternative baseline models.

Index Terms—Recommendation System, Cross-domain Recommendation, Variational Autoencoder, Deep Learning.

1 Introduction

The recommendation system is ubiquitous in modern life, covering various vital fields such as social networking, e-commerce, news, etc. It alleviates the problem of information overload by matching the relationships between users and products, which has attracted widespread attention from industry and academic fields. Conventional recommendation systems leverage collaborative filtering strategies to identify items a user is most likely to select from candidate groups. This technology follows the spirit that similar users will often prefer similar products and establishes user-item correlations. However, the interaction data in various real-world recommendation scenarios are generally very sparse. It is difficult to guarantee the data quantity and quality, making serving users with few or no interaction records intractable, thus limiting the model representation and the recommendation performance.

To tackle the above problems, cross-domain recommendation systems [2], [8] are proposed to mitigate data scarcity. For example, if a user likes the *Harry Potter* series, he will probably also prefer the *Harry Potter* movies adapted from the book. Particularly, as shown in Fig. 1, cross-domain recommendations (CDR) can be divided into single-target CDR and dual-target CDR through the difference in final goals. Previous works have focused on the former one [10], [28], [36] that transfers rich knowledge from the auxiliary domain (source domain) to boost the spares domain (target domain) recommendation accuracy. While the latter realizes the co-improvement in both domain performances by learning mutually. To make the most of product domain information, we focus on dual-target CDR. Under the CDR scenario, overlap users or products can be regarded as the explicit bridge to assist the model in identifying the implicit potential relationships. In real life, web platforms are split into diverse domains according to their service orientation, and people will inevitably interact with many of them to satisfy various needs. To improve the user experience in different domains, we explore cross-domain recommendations under the condition of sharing users.

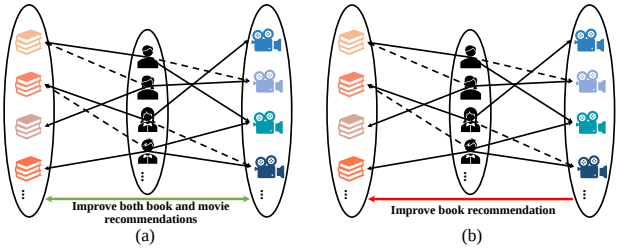

Fig. 1: Illustration of the shared user CDRs. (a) belongs to dual-target CDR. While (b) is categorized as single-target CDR. Solid arrows indicate the observed interactions. Dashed arrows represent missing interactions that need to be predicted.

Researchers have contributed a lot to CDR in different ways. [11] and [22] use the matrix factorization (MF) method, which evolved from collaborative filtering to learn

• *Tong Zhang, Chen Chen, Dan Wang, Jie Guo and Bin Song are with the State Key Laboratory of Integrated Services Networks, Xidian University, 710071, China. Bin Song and Dan Wang are the corresponding authors. E-mail: tongz@stu.xidian.edu.cn, chenchen_123@stu.xidian.edu.cn, danwang@xidian.edu.cn, jguo@xidian.edu.cn, bsong@mail.xidian.edu.cn.*

Manuscript received April 19, 2005; revised August 26, 2015.

the latent factors of sharing users. It simultaneously decomposes user matrices and embeds user and item features into the same low-dimensional space, ignoring that user preferences are not equally distributed across product domains. Therefore, it fails to capture complex interaction patterns. The emergence of algorithms based on deep neural networks (NN) alleviates the difficulty mentioned above to a certain extent. NN-based models enable the extraction of complicated patterns from interactions. Such as CoNet [10], it refines user features and fits similarities between domains with cross-connection unit. Despite the effectiveness, these methods are more concerned with refining the migration function, neglecting within-domain modeling or modeling within-domain preference features with insufficient precision, which may constrain transfer effectiveness when within-domain features predominate. Therefore, we model user behaviour features through both cross-domain and within-domain paths. Considering that a cross-domain recommendation scenario in which the user's interaction records in a domain may be very sparse. And the records are the reflection of user preference. In this context, learning a deterministic representation of a limited number of interaction products may result in information constraints (limited features to learn) and biases (uncertainty in user preferences). Therefore, we use the more generalisable VAE, which, in contrast to neural networks, does not focus on deterministic feature representations but on feature distributions. This property allows VAE to account for uncertainty in the latent space, which is particularly valuable when dealing with sparse data such as interaction records [3], [29]. Therefore, in this paper, we refine the within-domain user feature extraction by a deep Bayesian probability model, which combines Bayesian variational inference and neural network for the cross-domain recommendation. The Bayesian inference improves the accuracy of modeling user features within a domain, enabling more reliable and adequate knowledge migration. In addition to accurate modeling of user characteristics, the effective knowledge transfer between different domains is still a primary challenge. To better bridge the differences of preference across domains, many scholars have applied transfer learning to CDR and achieved performance improvement in recommendation. [14] adopts a dual transfer mechanism to mine bilateral related information on both the user side and item side, and [38] combines meta-learning and transfer-learning to construct a general framework for CDR. These methods need the training of many parameters and an elaborate network structure. For simplicity, an orthogonal transformation matrix is used for knowledge transfer. Its orthogonality allows better preservation of the inner product of latent vectors, thus learning similarities between different embeddings. At the same time, because the inverse matrix of an orthogonal matrix is equivalent to its transposed matrix, it can only realize knowledge transfer through simple operations.

Specifically, in this paper, we propose a framework for a dual-path modeling cross-domain recommendation based on VAE. It contains two parallel hybrid architecture VAEs as the backbone, guaranteeing generation ability and model robustness. The model is prone to falling into the posterior collapse problem for within-domain. That is, the model ultimately depends on the autoregressive characteristics of the decoder and ignores the influence of latent variables, which makes VAE valueless. We designed a complex prior expression and asymmetric codec structure to avoid this phenomenon, ensuring that the model has sufficient ability to mine domain-specific information. Then, for the cross-domain, the posterior knowledge conversion between domains is completed by training a transformation matrix. The converted information is fed into the decoder of the other domain to generate the user beheviour distribution.

Overall, we summarize the main contributions of our work as follows:

- We propose a VAE-based deep generatic Bayesian probability model for the dual-target cross-domain recommendation. It effectively eases the problem of sparse interactions and significantly enhances the performance of both domain recommendations simultaneously.
- We apply a dual-path modeling approach containing within-domain and cross-domain modeling. Domain-specific preferences are reconstructed through asymmetric codec structure and combined priors. Combined with the transformation matrix, the decoders learn auxiliary information and enhance the generating ability of cross-domain behaviour distributions.
- We conduct extensive experiments on amazon benchmark datasets to verify the superiority of CDRVAE. The outstanding results demonstrate our model reaches the state-of-the-art performance.

The rest of this paper is organized as follows: we first introduce the related work of cross-domain recommendation and variational autoencoder in section 2. In section 3, we formulate the problem definition. We propose the novel CDRVAE model and demonstrate the details in section 4. Experimental results and analysis of ablation experiments are given in section 5 to show the model's effectiveness. Finally, the conclusion of this paper and the discussion of future work are listed in section 6.

2 RELATED WORK

In this section, we give a brief review of related works, covering cross-domain recommendation and variational autoencoder.

2.1 Cross-domain Recommendation

Cross-domain recommendation emerges as a powerful way to solve the problem of sparse interaction records. In the early stage of research, studies concentrate on analyzing behaviour data to detect similar users to the target one and recommend the corresponding items, which are named nearest-neighbour-based methods [2]. These approaches are too simple to capture sufficient collaborative signals and are defeated by **MF-based models**. Through decoupling the joint user-item matrix of shared users in different domains, Collaborative Matrix Factorization (CFM) [27] exploits shared user factors. Hu et al. designed triadic factorization model CDTF [11] for a better relevance construction of users, items and domains. Loni et al. [19] treated the diverse categories of interactions as specific domains and

made use of auxiliary information via a factorization mechanism to achieve CDR. Considering potential commonalities in user groups, [22] applied K-means and SVD to factorize the cluster-level matrices. Man et al. [21] integrated MF into different domains, built a general mapping framework, and learned the function from the source domain to the target one. CCCFNet [15] composed MF and content-based filtering to construct a unified factorization network, which intends to find consistent patterns among domains. Although the above algorithms have made great progress in relieving data sparsity, they are inherently linear and cannot have the capability to model complex interactions. In addition, these approaches' performance greatly depends on the data. Under the cross-domain recommendation scenario, the user-item matrix, which is not dense originally, may become more sparse and lead to worse results.

With the development of deep learning, the application of neural networks has become a hot direction in CDR. Accompanied by the excellent nonlinear fitting capability, **NN-based models** are capable of capturing complicated interactive relationships in domains and constructing an effective projection function across domains. CoNet [10] transfers collaborative knowledge through the cross-mapping unit connecting domains. DARec [33] applies the domain adaptation methodology to seek domain-invariant transferable features. S. Ahangama et al. merged the latent user embeddings of dense and sparse domains to bridge domain-specific knowledge [6]. CDCFLFA [7] improves the accuracy of knowledge transfer by aligning the latent factors of the target domain and the auxiliary domain by pattern matching. Different from the unidirectional transfer in the above models, DTCDR [35] is the first proposed model for a dual-target cross-domain recommendation, which fuses and shares the embeddings from different domains according to multi-task learning. DDTCDR [14] introduces dual-transfer learning to achieve a two-way latent connection. Zhu et al. took into account the topology of the interaction relationship and applied the graph network to CDR [16]. The GADTCDR leverages a graphical and attentional framework for a more representative embedding extraction. These approaches investigate the methods of cross-domain knowledge transfer but neglect that the sparse interaction information may degrade the credibility and richness of knowledge. Graph networks can model the complicated interactions between users and products, which can learn higher-order embedding representations through information aggregation. Some studies apply graph neural networks to the scenario of cross-domain recommendation. BiTGCF [17] performs bidirectional knowledge transfer in both domains by fusing common and domain-specific features of shared users. A deep generative model mitigates this difficulty by modeling the preference distribution. It presents outstanding ability in data generation and feature representation, which also strengthens the explanation ability of inferencing. Salah et al. [24] utilized the variational inference to align the latent space to the source one. Although adequately modeling user preference characteristics within a domain, it directly transfers knowledge through latent variables, ignoring the heterogeneity between spaces. Compared with the above methods, our proposed method is a deep probabilistic model. Combining the advantages of deep neural networks and distribution generation, CDRVAE precisely excavates within-domain properties by a designated VAE and explores the cross-domain knowledge transfer scheme by training a transformation matrix.

Several efforts try applying other techniques to address the problems in the CDR. FedCT [18] reduces the demand for data by federated learning, avoids privacy leakage and improves system security. Zhu et al. [38] considered a transfer-meta framework to tackle the poor generalization ability to new domains and accomplish quick optimization for transfer modules. In this paper, we do not aim to cope with additional challenges. But in fact, our model is compatible with these meaningful approaches.

2.2 Variational Autoencoder

Variational autoencoders in recommendation have aroused great concern among scholars because of their potential in learning data distributions [26], [29]. CDAE [31] employs denoising autoencoders to form the feedback data and learn robust behaviour-distributed representations. Liang et al. [16] extended variational inference to collaborative filtering with implicit feedback. [26] modifies the regularization term to improve the model representation ability and optimizes the recommendation performance. BiVAE [29] explores a latent space from both the user and item sides and enhances the model's robustness. Although the above methods have successfully applied VAE in the recommendation, they only focus on a single domain. [24] proposes a VAE framework for CDR. However, it merely discusses single-target CDR scenarios.

We investigate the effectiveness of VAE in dual-target CDR. The model trains two sets of parallel encoders simultaneously so that the decoders can derive auxiliary latent representations.

Posterior collapse means that the posterior $q_\phi(z|x)$ contributes little information to the reconstruction when the KL term in Eq. 3 approaches 0 during the training. It may lead to disastrous results where the sample outputs of the decoder are separated from the true data distribution. To alleviate this issue, some methods decrease the regularization effect of KL. For instance, ControlVAE [25] multiplies the regularization term by a weight coefficient. Inspired by cybernetics, the controller is adopted to learn the weight automatically. [37] adds batch normalization operation to make KL has a lower bound greater than 0. In contrast, others concentrate on forcing VAE to accept latent factors by weakening the decoder. [4] randomly drops out some latent variables to make the decoder more dependent on the observed representations. [34] introduces additional loss, so the model has more concerns about reconstruction optimization.

In this work, to mitigate the influence of posterior collapse, we use combined priors to increase the difficulty of the KL constraint and design an asymmetric codec structure that contains a complicated encoder and a simple decoder to weaken the decoding ability.

3 PRELIMINARY

In this section, we first elaborate on CDR's detailed problem definition, then briefly introduce the variational autoencoders and some extensions.

3.1 Problem Definition

In this paper, we explore dual-target CDR under the shared-user scenario. This task aims to predict the items that users are most likely to interact with in each single domain based on the common user preference information from both domains. Different domains D_A and D_B share the same set of users, presented by $U = \{u_1, u_2, \ldots, u_n\}$, in which n is the total number of users. The items in D_A depicted as $I^A = \{i_1^A, i_2^A, \ldots, i_a^A\}$. While in D_B, the item set is $I^B = \{i_1^B, i_2^B, \ldots, i_b^B\}$. a and b denote the number of items in the corresponding domain, respectively. For domain D_A, the user-item preference distribution learnt from the interaction matrix $R^A \in \mathbb{R}^{n \times a}$, where each entry from explicit or implicit feedback is defined as a binary value:

$$r_{ui}^A = \begin{cases} 1, & \text{if } u \text{ positively interacts with } i \\ 0, & \text{otherwise .} \end{cases} \tag{1}$$

Similarly, the interaction matrix $R^B \in \mathbb{R}^{n \times b}$ for domain D_B.

3.2 Brief Review of VAE and Variants

VAE is a type of generative model. It feeds a set of real samples $x \in \mathcal{X}$ into an encoder network to map the data to a latent feature space $\mathcal{Z}$. x is the row of R, which can be taken as the interaction behaviour of one user for all items. Then, the obtained latent variables $z \in \mathcal{Z}$ are mapped back to the original distribution space to get the reconstructed samples through the decoder. By minimizing the reconstruction loss, VAE earns the ability to represent data accurately. The encoder is parameterized by θ. While ϕ is the parameter of the decoder. The training loss function is denoted by maximizing the marginal likelihood:

$$\mathbb{E}\left[\log p_\theta(x)\right] = \mathbb{E}\left[\log \mathbb{E}_{p(z)}\left[p_\theta(x|z)\right]\right]. \tag{2}$$

On account of the marginal likelihood, $p_\theta(x) = \int p_\theta(x|z)p(z)dz$ is intractable to integrate, Kingma et al. [13] defined a method named *amortized inference* to approximate the original optimization objective to the Evidence Lower BOund (ELBO).

$$\mathcal{L}_{\text{VAE}}(x) = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z) - D_{\text{KL}}\left(q_\phi(z|x)\|p(z)\right)\right]. \tag{3}$$

This formulation consists of a reconstruction error and a regularization term. In Eq. 3, D_{KL} is the Kullback-Leibler divergence which measures the similarity between distributions. $p(z)$ presents the prior distribution of data. $q_\phi(z|x)$ denotes the approximate posterior. $p_\theta(x|z)$ implies a conditional distribution.

Denoising variational autoencoders (DVAE) [30] adds noise to the input. Then, it reconstructs the uncorrupted data from damaged samples. The encoder has learned to extract the crucial features and exclude the interferences. DVAE is more robust and less prone to overfitting. Its ELBO is formulated as follows:

$$\begin{aligned} \mathcal{L}_{\text{DVAE}}(x) = \mathbb{E}_{q_\phi(z|\tilde{x})}\mathbb{E}_{q_\phi(\tilde{x}|x)}[\log p_\theta(x|z) \\ - D_{\text{KL}}\left(q_\phi(z|\tilde{x})\|p(z)\right)]. \end{aligned} \tag{4}$$

Mult-VAE assumes the prior follows the standard Gaussian distribution and selects multinomial distribution as the likelihood function. A softmax operation, followed by a multi-layer perceptron, computes the probability vectors. Given the total number n of users, user behaviour can be considered sampled from the multinomial distribution.

$$\begin{aligned} z \sim \mathcal{N}(0, \mathrm{I}), \quad \pi(z) = \operatorname{softmax}(f(z)) \\ x \sim \operatorname{Mult}(n, \pi(z)). \end{aligned} \tag{5}$$

The ELBO takes the following form:

$$\begin{aligned} \mathcal{L}_{\text{Mult-VAE}} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x \mid z) \\ - \beta \mathrm{D}_{\text{KL}}\left(q_\phi(z \mid x) \| p(z)\right)], \end{aligned} \tag{6}$$

where $q(z)$ is assumed sampling from a fully factorized (diagonal) Gaussian distribution: $q(z) = \mathcal{N}(\mu, \operatorname{diag}(\sigma^2))$. The weight coefficient β enables the balance between representation ability and reconstruction ability determined by annealing.

Fig. 2 illustrates the structure of all above-mentioned VAE networks.

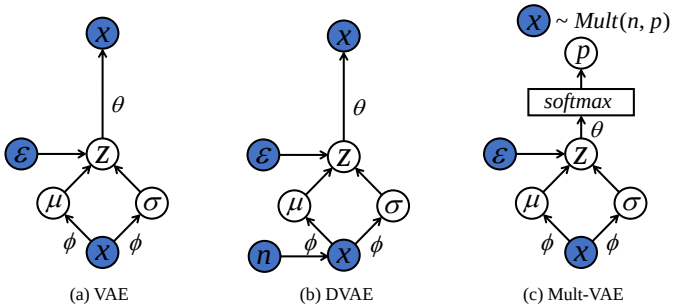

Fig. 2: VAE and its variants.

DVAE enhances robustness by rebuilding knowledge from disturbing information. Mult-VAE refines user feature modeling by multinomial distribution. Our approach is based on DVAE and Mult-VAE, inheriting their respective advantages. The detailed modifications are described in Section 4.1.

4 PROPOSED METHOD

In this section, we propose a new advanced VAE-based deep Bayesian generative model for the dual-target cross-domain recommendation, CDRVAE. For the given user-item interaction matrix, the user preference features within a domain follow two streams oriented towards two domains, which we call dual-path modeling. Take the domain D_A for a detailed explanation. One stream flows from the encoder Enc_A to the decoder Dec_A. The elaborate encoder Enc_A first learns a low-dimensional latent space from partially observed records to obtain the posterior knowledge so that the decoder can reconstruct the unobserved interactions. Another flow flows from Enc_A to Dec_B. The latent factors of D_A are converted by a transformation matrix which captures a transfer relationship across domains to assist D_B in learning relevant preference features of common users. The overall architecture of the proposed model is depicted in Fig. 3.

4.1 Hybrid Architecture

As discussed in Section 3.2, Mult-VAE and DVAE have made incredible achievements in representativeness and

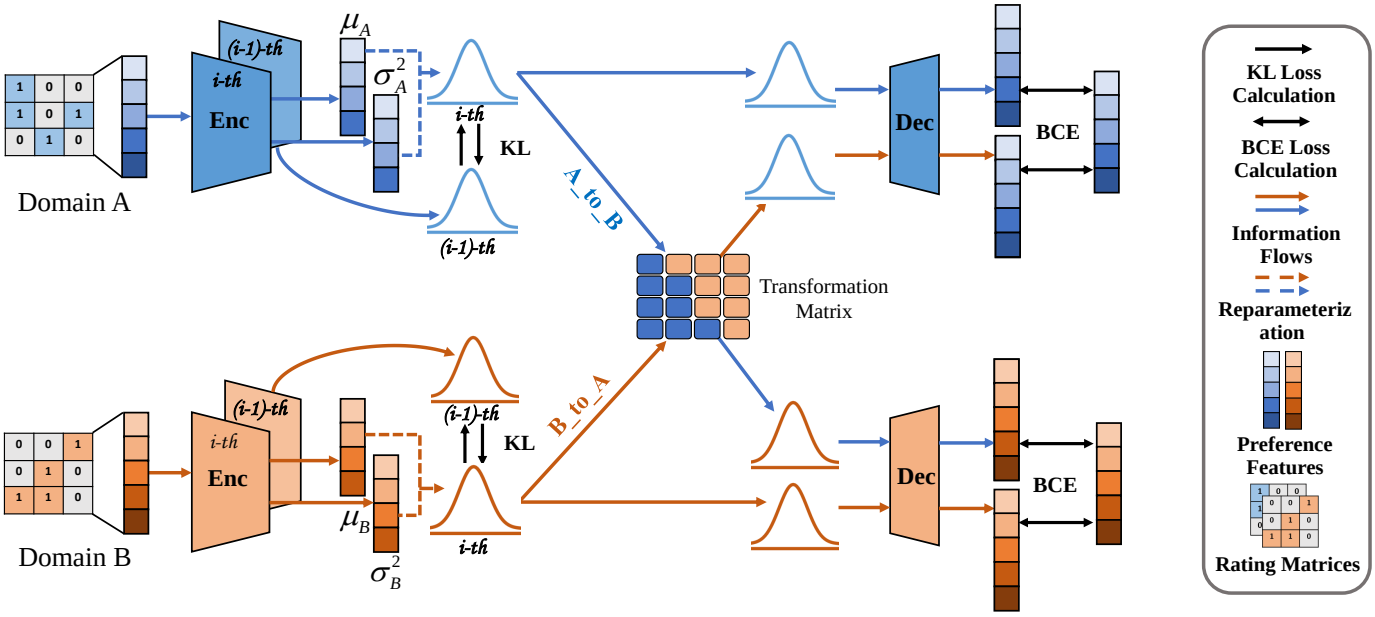

Fig. 3: An illustration of the architecture of CDRVAE. For within-domain modeling, two parallel sets of VAE explore latent spaces of preferences in two domains and learn the domain-related distributions by reconstructing the original data. For cross-domain modeling, a transformation matrix learns an equivalent conversion of preference between heterogeneous domains, and the decoders regenerate the user behaviour from converted latent factors. Due to the same user's preferences in different domains being intrinsically related, within and cross-domain modeling approaches facilitate the CDR.

robustness. We combine them to design a hybrid structure, which serves as the backbone for feature extraction. The latent space is assumed to be a standard isotropic Gaussian distribution. The corrupted $\tilde{x}$ follows a multinominal distribution conditioned on the latent variables. The formulations of the hybrid structure can be rewritten as follows:

$$q_\phi\left(z|\tilde{x}\right) = \mathcal{N}\left(z|f_\phi\left(\tilde{x}\right)\right), \tag{7}$$

$$\pi\left(z\right) = \mathrm{softmax}\left(f_\theta\left(z\right)\right), \tag{8}$$

$$p_\theta\left(x|z\right) = \mathrm{Mult}\left(x|n, \pi\left(z\right)\right). \tag{9}$$

The detailed theoretical analysis refers [16], [30]. $\tilde{x}$ denotes the corrupted input data in which some interactions are randomly discarded before entering the model. $f_\phi(\cdot)$ is an amortized inference network defined by ϕ. To approximate the real posterior $p(z|x)$, the inference network returns the mean vector μ_ϕ and variance matrix σ_ϕ of latent factor z. The reparametrization trick establishes the continuous latent space by using output values.

$$z = \mu_\phi(\tilde{x}) + \varepsilon \odot \sigma_\phi(\tilde{x}), \tag{10}$$

where ε is random noise sampled from a standard Gaussian. After the decoding process by $f_\theta(\cdot)$ and a softmax operation, a probability vector over whole products is obtained, where each element of $\pi(z)$ represents the probability that the user selects the corresponding item. f_ϕ and f_θ are implemented by neural networks. It's notable that they are independent of each other and have different structures. The structural details are interpreted in Section 4.2.

As a result, we obtain a hybrid VAE architecture that possesses the advantages of precisely describing user preferences through multinomial distribution and accurately reconstructing behaviour distribution from corrupted data. This architecture will be applied in both D_A and D_B for sufficient feature extraction.

Intuitively, the behaviours generated by the same user in two domains are intrinsically related. Establishing cross-domain reconstruction is critical for CDR to help mine consistent information from the heterogeneous domain. Therefore, CDRVAE explores more indicative interaction preferences by knowledge transformation, which assists in predicting absent behaviours. Referring to the loss function of DVAE and Mult-VAE in Section 3.2. The ultimate objective function for CDR is formulated as follows:

$$\begin{aligned}\mathcal{L} =& \mathbb{E}_{q_\phi(z_A,z_B|\tilde{x}_A,\tilde{x}_B)}\mathbb{E}_{p(\tilde{x}_A,\tilde{x}_B|x_A,x_B)}\left[\log p_\theta\left(x_A,x_B|z_A,z_B\right)\right. \\ &\left.-\beta\mathrm{KL}\left(q_\phi\left(z_A,z_B|\tilde{x}_A,\tilde{x}_B\right)\|p\left(z_A,z_B\right)\right)\right],\end{aligned} \tag{11}$$

where x_A and x_B denote the user behaviours from D_A and D_B. All subscripts $*_A/*_B$ indicate the domain to which the data belongs. $\tilde{x}_A$ and $\tilde{x}_B$ are corrupted data where part of them are masked randomly. z_A and z_B correspond to latent variables encoded by amortized inference networks.

To make the optimization objective easier to calculate, we assume that user behaviours in different domains are interrelated and can be reconstructed from the transferred information. This can be intuitively understood that a user's preference will affect his interest in different domains. Since the transformation matrix converts the knowledge signals equivalently, the following propositions can be concluded: when the observed interactions are given, the latent representations are conditional independent. When the latent spaces are determined, the reconstruction results are conditional independent.

According to whether the information flows to the other

domain, the above objective function is divided into two parts: within-domain loss and cross-domain loss.

4.2 Within-domain Modeling

Accurate modeling of user preferences within a domain is indispensable not only for improving the performance of its domain, but also significant for obtaining sufficient ancillary information in the other domain, which was underlooked in previous works.

Asymmetric Codec Structure. A strong autoregressive decoder may lead to a posterior collapse problem. To avoid converging to a degenerated local optimum, an asymmetric codec structure is designed to force the weakened decoder to accept more latent variables. Both domains possess the same network frame, and the details of one side are described below. Notably, the parameters of VAE on both sides are not shared.

Inspired by the information bottleneck theory, the encoder learns as many domain attribute features as possible. At the same time, the decoder decodes the most relevant generic features from it. For simplicity, a single fully connected layer is applied as the decoder.

$$f_\theta(z) = Wz + b. \tag{12}$$

Because the decoder's linearity restricts its capacity to reconstruct the original distribution from the latent variables, the decoder needs to choose more meaningful information from the approximate posterior.

On the contrary, the inference network is implemented by a non-linear neural network. It consists of five non-linear mapping layers. For each hidden layer, the tanh activation function introduces non-linearities, and a layer-normalized operation is employed to standardize the input. Inspired by [37], which stems the KL vanishing by maintaining its expectation positive, our encoder adds an additional batch normalization layer after the output of mean vectors, aiming to converse the μ_ϕ to a distribution with a fixed mean and variance. The final outputs are calculated from the following formulations:

$$\begin{aligned} \mu_\phi^{'}, \sigma_\phi^{'} &= f_\phi(\tilde{x}), \\ \mu_\phi = BN\left(\mu_\phi^{'}\right), &\quad \sigma_\phi = \sigma_\phi^{'}, \end{aligned} \tag{13}$$

where $BN(\cdot)$ indicates the batch normalization layer.

Combined Prior Strategy. With the continuous updation of encoder parameters, the approximate posterior gradually approaches the hypothesized true prior. Generally, the unknown prior is assumed to be a standard normal distribution. However, recent works [1], [32] tend to consider a complex prior to characterizing the latent factor. The goal of amortized inference is to expect the posterior to approach the prior. So we propose allowing part of the posterior information as supplementary information to the prior. The combined prior can be interpreted as a union of theoretical assumptions and practical approximations. Specifically, the approximate posterior from the previous iteration is added. For the $i-th$ step training, the combined prior is drawn from the following formula:

$$p(z|\phi_{\text{i-1}}, x) = \lambda\mathcal{N}(z|0, \mathrm{I}) + (1-\lambda)q_{\phi_{\text{i-1}}}(z|x). \tag{14}$$

The sum of the standard Gaussian distribution and the previous posterior defined by the parameters from the $(i-1)-th$ iteration is taken as the combined prior. The prior conditions of $\phi_{\text{i-1}}$ and x, and λ is an adjustable weight factor. The value of λ is set as 0.25 to allow the model to inherit more information from the previous training results. For an alternative interpretation, the second term brings the prior closer to the approximate posterior and regulates the training stride. During the optimization, the combined prior will prevent the miss of the globally optimal solution.

Loss Function. The loss function of reconstructing within domain information flow is as follows:

$$\begin{aligned} \mathcal{L}_{wd} = &\mathbb{E}_{q_{\phi_A}(z_A|\tilde{x}_A)}\mathbb{E}_{p(\tilde{x}_A|x_A)}[\log p_{\theta_A}(x_A|z_A) \\ &- \beta\mathrm{KL}\left(q_\phi(z_A|\tilde{x}_A) \,\|\, p(z_A)\right)] \\ &+ \mathbb{E}_{q_{\phi_B}(z_B|\tilde{x}_B)}\mathbb{E}_{p(\tilde{x}_B|x_B)}[\log p_{\theta_B}(x_B|z_B) \\ &- \beta\mathrm{KL}\left(q_\phi(z_B|\tilde{x}_B) \,\|\, p(z_B)\right)]. \end{aligned} \tag{15}$$

The weight β is set to a number significantly less than 1.

4.3 Cross-domain Modeling

We assume that two domains share the same set of users. Their tastes in a certain type of product are intrinsically related. Therefore, it is essential to utilize cross-domain knowledge to understand behaviour distributions better.

Transformation Matrix. The premise of this task is shared users, so a user's preference in one domain affects his behaviour in the other domain. Correspondingly, the mapping function should preserve the source domain attributes and represent effective conversion relations. Referring to the definition in mathematics, the transpose of an orthogonal matrix is equal to its inverse, so the latent factors can be easily recovered, such as $z_A = z_A W^T W$, which constrains the transformation matrix to retain the domain features. Besides, the cross-domain information transfer is accomplished by a shared matrix, so value of the matrix is constrained by features from two domains. During the convergence process, the vector inner product preserves the commonalities across multiple latent factors, which is beneficial for inter-domain shared feature learning. Thus, a trainable transformation matrix is employed as a bridge for cross-domain knowledge transfer. To ensure the orthogonality of the transformation matrix, the original latent factors and the recovery version are drawn closer to each other by using the L_2 norm.

Cross-domain Reconstruction. The cross-domain reconstruction of converted features affects the transfer optimisation and influences the final recommendation results. The latent feature factor $\overline{z_A}$, which is calculated by $z_B W$ is passed into the domain D_B, denoting the auxiliary information from D_A. All latent variables represent user behaviour distribution. The combination of $\overline{z_A}$ and z_B will cause a new distribution to deviate from the user's real preference and introduce some noise interference, affecting the reconstruction of user features. If applying an additional decoder will bring in more parameters and make model convergence difficult. Therefore, in CDRVAE, $\overline{z_A}$ is directly fed into Dec_B like z_B to generate the D_B preference distribution. $\overline{z_B}$ follows the same procedure. In this way, the transfer process avoids the impact of noise generated by the poor performance of the transformation matrix at the beginning

of training by reconstruction loss, and CDRVAE also reduces the network parameters. Thus, the decoder is forced to selectively accept information. In other words, the preference feature needs to be recreated simultaneously from the approximate posteriors of two versions. The decoder can not only have the ability to reconstruct the interaction distribution from within-domain latent factors, but also generate missing behaviours from the cross-domain transfer information.

Loss Function. The loss function of reconstructing cross-domain information flow is as follows:

$$\begin{aligned}\mathcal{L}_{cd} = & \mathbb{E}_{q_{\phi_A}(z_A|\tilde{x}_A)}\mathbb{E}_{p(\tilde{x}_A|x_A)}[\log p_{\theta_B}(x_B|\overline{z_A}) \\ & - \beta \mathrm{KL}\left(q_\phi\left(z_B|\tilde{x}_A\right)\|p\left(z_A\right)\right)] \\ & + \mathbb{E}_{q_{\phi_B}(z_B|\tilde{x}_B)}\mathbb{E}_{p(\tilde{x}_B|x_B)}[\log p_{\theta_A}(x_A|\overline{z_B}) \\ & - \beta \mathrm{KL}\left(q_\phi\left(z_A|\tilde{x}_B\right)\|p\left(z_B\right)\right)] \\ & + \gamma\left(\|z_A - z_A W^T W\|_F^2 + \|z_B - z_B W W^T\|_F^2\right),\end{aligned} \tag{16}$$

where $\overline{z_A}$ and $\overline{z_B}$ denote the auxiliary information calculated by $z_B W$ and $z_A W^T$, respectively. $\|\cdot\|_F^2$ is the L_2 regularization. The prior $p(z)$ and β have the same definition as within-domain modeling.

4.4 Total Loss

Since CDRVAE focuses on implicit feedback, we apply Binary Cross Entropy Loss to measure the reconstruction quality. Combing within-domain and cross-domain paths, the model is optimized by the final loss:

$$\mathcal{L} = \mathcal{L}_{wd} + \eta \mathcal{L}_{cd}. \tag{17}$$

η is a trade-off factor in the range 0 to 1. Considering that transfer signals play an auxiliary role in CDR, we set the value as 0.2. Therefore, the proposed model can acquire both the within-domain characteristics and the cross-domain supplementary information.

5 EXPERIMENTS

To validate the effectiveness of our proposed model, we compare the cross-domain recommendation accuracy with several state-of-the-art methodologies. Furthermore, we conduct extensive analysis to answer the following questions:

- Does CDRVAE outperform present state-of-the-art models in benchmark datasets?
- Does every improvement is beneficial to the final performance?
- How the different VAE architectures affect CDRVAE?
- How does CDRVAE perform with different transfer factors of cross-domain knowledge?
- How do the complexities of the encoder and decoder affect the experimental results?
- Whether the model performs well on datasets with sparsity differences and overlapping user scales?

5.1 Datasets and Preprocessing

We verify the superiority of CDRVAE on three anonymous public sub-datasets from Amazon, which consist of user explicit rating feedback and has been commonly applied in cross-domain literature, covering Books (**Books**), Movies and TV (**Movies**) and CDs and Vinyl (**Music**). Following [5], we treat users with rating records greater than 5 as valid users and define the ratings of 3-5 as positive samples and the rest as negative. To form cross-domain datasets, we combine the three datasets into pairs and select shared users and their interaction records in paired domains. Finally, **Movies&Books**, **Movies&Music** and **Books&Music** are acquired. All datasets are highly sparse, with a maximum 0.14% interaction density. The detailed statistics are summarized in TABLE 1.

TABLE 1: Descriptive statistics of the used datasets.

Datasets	Movies&Books		Movies&Music		Books&Music	
Users	29,476		15,914		16,267	
Domain	Movies	Books	Movies	Music	Books	Music
Items	24,091	41,884	17,794	20,058	23,988	18,467
Ratings	591,258	579,131	416,228	280,398	291,325	233,251
Density	0.08%	0.05%	0.14%	0.09%	0.07%	0.08%

5.2 Evaluation Metrics

The leave-one-out (LOO) assessment is frequently employed [10] in recommendation literature, and we also apply it for model training. Specifically, one random interaction is allocated as a test item for each user, and the rest is reserved as the training set. Following previous works [5], [10], a random selection of 99 negative samples with unobserved interaction and one positive item are leveraged to assess CDRVAE's ranking results of test items against negative entries. We adopt the commonly used evaluation criteria for performance quantification: HR, NDCG and MRR. We truncate the returned rank list for all metrics at 5 and 10. Higher values of them give better predicting performance.

5.3 Baseline Models

We compare CDRVAE with several single-domain and cross-domain baseline models:

PMF [23] is a classic probability algorithm for single-domain recommendation, which learns user and item features by matrix factorization.

AAE [20] is a hybrid of GAN and VAE. We modified it for a single-domain recommendation task.

CDAE [31] is designed for single-domain recommendation. It adds an additional user node with specific weights in the input layer and a bias node in the hidden layer for feature extraction.

CFVAE [16] is a VAE-based algorithm for collaborative filtering with insufficient training samples.

CMF [27] is a shallow latent factor model, which jointly factorizes the matrices of multiple domains.

AAE++ [20] is an extension of AEE. It applies an adversarial autoencoder to distinct domains and makes performance improvements.

TABLE 2: Performance comparison on Movies&Books. The best results are lit in bold, and the second best are marked underlined. Compared to the corresponding best baseline, † denotes p-value<0.01, ‡ denotes p-value<0.05 and - denotes p-value>0.05.

Domain	Movies			Books			Movies			Books		
Metrics	HR@5	NDCG@5	MRR@5	HR@5	NDCG@5	MRR@5	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
PMF	0.4664	0.3147	0.2745	0.4003	0.2961	0.2621	0.5737	0.3591	0.2928	0.5121	0.3327	0.2772
CDAE	0.4660	0.3471	0.3056	0.4483	0.3492	0.3157	0.5991	0.3901	0.3263	0.5640	0.3851	0.3315
CFVAE	0.4587	0.3396	0.3006	0.4277	0.3258	0.2918	0.5928	0.3852	0.3206	0.5508	0.3646	0.3091
AAE	0.4661	0.3471	0.3080	0.4457	0.3509	0.3128	0.5989	0.3900	0.3269	0.5559	0.3871	0.3591
CMF	0.4433	0.3224	0.2815	0.4373	0.3225	0.2848	0.5848	0.3674	0.3000	0.5583	0.3616	0.3009
AAE++	0.4803	0.3590	0.3189	0.4537	0.3592	0.3280	0.6098	0.4009	0.3362	0.5656	0.3954	0.3429
CoNet	0.3886	0.2702	0.3379	0.3451	0.2361	0.3108	0.5244	0.3145	0.2970	0.4690	0.2716	0.2653
DDTCDR	0.4090	0.2942	0.2576	0.4008	0.3153	0.2893	0.5394	0.3382	0.2732	0.5073	0.3492	0.3013
DARec	0.4914	0.3641	0.3224	0.4690	0.3591	0.3227	0.6202	0.4069	0.3401	0.5919	0.3989	0.3392
ETL	0.5115	0.3812	0.3431	0.5111	0.3989	0.3705	0.6419	0.4244	0.3608	0.6329	0.4383	0.3861
Bi-TGCF	0.5409	0.4039	0.3633	0.4718	0.3590	0.3222	0.6695	0.4455	0.3802	0.5884	0.3967	0.3373
CDRVAE	**0.5846**†	**0.4518**†	**0.4085**†	**0.5252**†	**0.4088**†	**0.3708**$^{-}$	**0.7104**†	**0.4629**†	**0.4254**†	**0.6518**†	**0.4496**†	**0.3871**†
Improv.	**8.1%**	**11.9%**	**12.4%**	**2.8%**	**2.7%**	**0.3%**	**6.1%**	**3.9%**	**11.9%**	**3.0%**	**2.6%**	**0.3%**

TABLE 3: Performance comparison on Movies&Music. The best results are lit in bold, and the second best are marked underlined. Compared to the corresponding best baseline, † denotes p-value<0.01, ‡ denotes p-value<0.05 and - denotes p-value>0.05.

Domain	Movies			Music			Movies			Music		
Metrics	HR@5	NDCG@5	MRR@5	HR@5	NDCG@5	MRR@5	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
PMF	0.4081	0.2872	0.2474	0.4505	0.3350	0.2969	0.5490	0.3326	0.2261	0.5769	0.3759	0.3137
CDAE	0.4191	0.3093	0.2723	0.4433	0.3396	0.3053	0.5544	0.3528	0.2898	0.5662	0.3792	0.3225
CFVAE	0.4318	0.3110	0.2750	0.4362	0.3281	0.2884	0.5699	0.3605	0.2945	0.5646	0.3663	0.3082
AAE	0.4357	0.3226	0.2860	0.4557	0.3445	0.3086	0.5689	0.3658	0.3023	0.5772	0.3863	0.3248
CMF	0.4309	0.3025	0.2603	0.4794	0.3568	0.3166	0.5736	0.3487	0.2793	0.6124	0.4011	0.3349
AAE++	0.4281	0.3142	0.2754	0.4538	0.3501	0.3142	0.5628	0.3564	0.2928	0.5789	0.3887	0.3301
CoNet	0.3729	0.2556	0.3183	0.3887	0.2658	0.3273	0.5146	0.3013	0.2759	0.5380	0.3140	0.2873
DDTCDR	0.3880	0.2748	0.2366	0.4204	0.3169	0.2804	0.5220	0.3177	0.2542	0.5421	0.3563	0.2962
DARec	0.4589	0.3349	0.2950	0.4822	0.3636	0.3241	0.5973	0.3790	0.3134	0.6125	0.4051	0.3413
ETL	0.4891	0.3632	0.3224	0.5314	0.4037	0.3653	0.6241	0.4076	0.3404	0.6550	0.4442	0.3819
Bi-TGCF	0.5064	0.3710	0.3262	0.5313	0.3978	0.3260	0.6401	0.4142	0.3441	0.6557	0.4382	0.3440
CDRVAE	**0.5125**‡	**0.3830**†	**0.3402**†	**0.5362**‡	**0.4138**‡	**0.3726**‡	**0.6452**$^{-}$	**0.4259**†	**0.3579**†	**0.6626**$^{-}$	**0.4533**‡	**0.3889**†
Improv.	**1.2%**	**3.2%**	**4.3%**	**0.9%**	**2.5%**	**2.0%**	**0.8%**	**2.8%**	**4.0%**	**1.1%**	**2.0%**	**1.8%**

TABLE 4: Performance comparison on Books&Music. The best results are lit in bold, and the second best are marked underlined. Compared to the corresponding best baseline, † denotes p-value<0.01, ‡ denotes p-value<0.05 and - denotes p-value>0.05.

Domain	Books			Music			Books			Music		
Metrics	HR@5	NDCG@5	MRR@5	HR@5	NDCG@5	MRR@5	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
PMF	0.4015	0.3182	0.2889	0.4213	0.3138	0.2783	0.4992	0.3480	0.3009	0.5360	0.3508	0.2936
CDAE	0.4046	0.3129	0.2868	0.4266	0.3259	0.2839	0.5139	0.3478	0.2985	0.5471	0.3615	0.3031
CFVAE	0.3763	0.2891	0.2891	0.4101	0.3104	0.2718	0.5077	0.3275	0.2747	0.5342	0.5342	0.2860
AAE	0.3983	0.3159	0.2852	0.4302	0.3326	0.3007	0.5121	0.3491	0.2992	0.5498	0.3712	0.3152
CMF	0.4017	0.3126	0.2920	0.4113	0.3084	0.2748	0.5132	0.3468	0.3055	0.5280	0.3468	0.2906
AAE++	0.3996	0.3200	0.2917	0.4270	0.3287	0.2956	0.5084	0.3535	0.3055	0.5450	0.3661	0.3110
CoNet	0.3265	0.2032	0.2899	0.3380	0.2235	0.2963	0.4505	0.2452	0.2419	0.4699	0.2663	0.2506
DDTCDR	0.3689	0.2992	0.2734	0.3965	0.3061	0.2749	0.4700	0.3300	0.2872	0.5110	0.3412	0.2879
DARec	0.4368	0.3350	0.3013	0.4535	0.3422	0.3060	0.5494	0.3710	0.3161	0.5796	0.3832	0.3229
ETL	0.4496	0.3493	0.3155	0.4686	0.3683	0.3282	0.5669	0.3865	0.3369	0.5942	0.4034	0.3444
Bi-TGCF	0.4044	0.3032	0.2701	0.4729	0.3620	0.3216	0.5189	0.3406	0.2854	0.6085	0.4022	0.3382
CDRVAE	**0.4733**†	**0.3601**‡	**0.3270**†	**0.4862**‡	**0.3765**†	**0.3449**†	**0.5891**†	**0.3975**†	**0.3423**‡	**0.6095**‡	**0.4161**†	**0.3527**†
Improv.	**5.3%**	**3.1%**	**3.6%**	**2.8%**	**2.2%**	**5.1%**	**3.9%**	**2.8%**	**1.6%**	**0.2%**	**3.1%**	**2.4%**

CoNet [10] introduces a cross-stitch module to transfer knowledge across domains. It explores the mapping relations of different domain features and jointly optimizes both domains.
DARec [33] adopts a domain adaptation technique for CDR. It learns preference patterns from interaction matrices.
DDTCDR [14] designs a dual-transfer learning mechanism for learning additional information. In each domain, we decompose the rating matrices to obtain user and item properties.
ETL [5] proposes an equivalent transformation to excavate user preferences across domains and uses joint distribution to explore domain relevance.
Bi-TGCF [17] integrates the common user and domain-specific features in the knowledge transfer process and uses the overlapping user as a bridge to realize the bi-directional transfer.

5.4 Implement Details

We fix the dimension of the latent vector as 200, which remains constant throughout training on both domains. The total number of epochs is set as 500 to guarantee sufficient convergence of all models. All parameters of CDRVAE are randomly initialized by Xavier initializer [9] and optimized by Adam optimizer [12]. The learning rate for the entire training process is set as 0.001, and the message dropout ratio is 0.5. Moreover, for the hyperparameter λ in the combined prior, we set it as $1/4$. Since the additional coefficient β in the regularization term of ELBO affects the final results, for each dataset, we search for the appropriate value in the range of 0.001 to 0.005. Finally, 0.005 was used for Movies&Books and 0.001 for Movies&Music and Books&Music. For the coefficient γ of the regularization term, 5.0 for Movies&Books, 0.5 for Movies&Music and 1.0 for Books&Music. We fixed the η as 0.2 in all experiments. To ensure the reasonableness of the experiments, we tune the knowledge transfer coefficient in the baseline models to match the paired datasets and select better results from them.

Our CDRVAE model is implemented by PyTorch framework and runs on hardware equipment with Intel Core i7-10700k 8-core 3.80GHz CPU, NVIDIA GeForce RTX 3090 GPU.

5.5 Overall Comparison (RQ1)

TABLE 2, TABLE 3, and TABLE 4 reflect the experimental results of CDRVAE and other baselines on three different datasets, respectively. Our model significantly and consistently yields outstanding performance in terms of the metrics with HR, NDCG and MRR through all paired datasets, which confirms the validity of CDRVAE.

In particular, CDRVAE outperforms all single-domain recommendation methods, demonstrating that cross-domain knowledge transfer can effectively mitigate data sparsity. Notably, CoNet and DDTCDR are inferior to other CDR algorithms because in the experiment setting, only interaction records are utilized in the training process, and no additional user or item features are offered as in the original paper. This phenomenon indicates that our model can learn reasonable user preference patterns only through the history records, without requiring private information such as age, gender, occupation and other sensitive information or product-related attributes. CDRVAE beats the competitive model DARec, which explores behaviour patterns via domain adaptation techniques. This may be because our model not only enables knowledge transfer but also emphasizes accurate modeling of within-domain preference. In Movies&Books dataset, CDRVAE achieves considerable performance improvements on top5 and top10 in terms of MRR by 12.4% and 11.9% on the movie domain compared with Bi-TGCF, indicating that the within-domain modeling is useful. In the book domain, the sparser one, CDRVAE also achieves performance gains over ETL. The superiority may be attributed to the dual-path modeling algorithm. The hybrid VAE architecture and the complex combined prior are capable of fitting complicated behaviour distribution of users. So the inner domain features are more representative and more beneficial knowledge transformed by cross-domain modeling.

Additionally, compared with Movies&Music and Books&Music, CDRVAE achieves a more substantial improvement in dual-domain recommendation performance on Movies&Books. This phenomenon is possible because books and movies have a stronger potential correlation, allowing for more useful secondary information for user performance prediction. In addition, it can be observed from TABLE 1 that Movies&Books has more rating records than the other two, from which more behaviour information is mined to support the modeling.

Moreover, on Movies&Books and Movies&Music, a relatively large difference appears in the recommendation performance improvement of the sparse and dense domains. The dense interactive domain even has a considerable boost over the sparse one. This may be because these two datasets have a more noticeable interaction density difference than the Books&Music. Cross-domain knowledge transfer will provide more auxiliary information to improve accuracy to a certain extent. Still, there may be negative transfer during the conversion process that affects the learning of the decoder. The negative transfer is more likely to limit performance gains in the sparse domain than the dense one.

5.6 Ablation Study (RQ2)

To illustrate that each component used in CDRVAE does contribute to performance improvement. We have conducted a series of ablation experiments.

We conduct experiments on three datasets and elect HR@10 as the evaluation metric. The series of the ablation study results are summarised in TABLE **??**. A checkmark indicates that CDRVAE applies the corresponding component. We can observe that every element in CDRVAE improves the final results. Using all components simultaneously, CDRVAE achieves optimal results. We set the decoder for the asymmetric codec structure to become the same as the encoder. Thus, the codecs of the model are converted into symmetric. As for the combined prior strategy, we replace it with a standard Gaussian distribution during our experiments. The two-layer non-linear neural network is used as a replacement for the transformation matrix. The asymmetric codec structure and the combined prior strategy

simultaneously mitigate the posterior collapse problem to ensure accurate modeling of behaviour distribution of users in the domain. From the experimental results, both of them boost performance improvement together. Employing only one of them cannot achieve better results. It is also worth noting that the final performance deteriorates when a non-linear network is employed for knowledge transfer. The poor ability of the non-linear network to retain features in the domain may cause some beneficial information loss.

Asymmetric Codec Structure		✓			✓	✓		✓
Combined Piror Strategy			✓		✓		✓	✓
Transformation Matrix				✓		✓	✓	✓
Movies	0.6607	0.7038	0.7034	0.7046	0.7064	0.7037	0.7077	**0.7104**
Books	0.6156	0.6462	0.6436	0.6464	0.6482	0.6490	0.6408	**0.6518**
Movies	0.6093	0.6382	0.6401	0.6396	0.6414	0.6345	0.6416	**0.6452**
Music	0.6110	0.6568	0.6517	0.6568	0.6617	0.6556	0.6600	**0.6626**
Books	0.5481	0.5848	0.5822	0.5846	0.5861	0.5878	0.5879	**0.5891**
Music	0.5755	0.6046	0.6006	0.6032	0.6059	0.6056	0.5923	**0.6095**

5.7 Impact of Hybrid VAE Architecture (RQ3)

In this part, we modify the backbone of CDRVAE to verify the effectiveness of hybrid VAE architecture. We conduct comparative experiments by replacing it with vanilla VAE, DVAE and Mult-VAE, denoted as CDRVAE-V, CDRVAE-D and CDRVAE-M, respectively. The inference network modifies the output layer to be suitable for variants. For fairness, the other settings of the model remain the same. We test on three datasets and calculate metrics based on the first ten returned items. The detailed experimental results are in the following three tables.

TABLE 5: Performance comparison $w.r.t$ different VAE architecture on Movies&Books.

Methods	Books			Movies		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
CDRVAE	**0.7104**	**0.4629**	**0.4254**	**0.6518**	**0.4496**	**0.3871**
CDRVAE-V	0.5907	0.3808	0.3804	0.5505	0.3622	0.3005
CDRVAE-D	0.6361	0.4194	0.3511	0.6285	0.4200	0.3457
CDRVAE-M	0.6861	0.4702	0.4040	0.6335	0.4305	0.3687

TABLE 6: Performance comparison $w.r.t$ different VAE architecture on Books&Music.

Methods	Movies			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
CDRVAE	**0.6452**	**0.4259**	**0.3579**	**0.6626**	**0.4533**	**0.3889**
CDRVAE-V	0.5645	0.3584	0.2925	0.5586	0.3601	0.2970
CDRVAE-D	0.6139	0.4062	0.3387	0.6438	0.4301	0.3714
CDRVAE-M	0.6415	0.4192	0.3531	0.6523	0.4460	0.3865

We choose the Books&Music for analysis. As shown in TABLE 5, CDRVAE with hybrid architecture performs better than DVAE-based and Mult-VAE-based models. The cross-domain framework with the vallina VAE shows the worst performance. Modeling interests are oversimplified when applying vanilla VAE. It doesn't consider the diversities in user preferences. There is also no operation such as noise reduction like DAVE to improve the robustness of the model, resulting in poor generalization ability in different domains. Although CDRVAE-D introduces noise to enhance preference representation, it gives the second worst experimental results among the three models. The reason may be that multinomial likelihood, which has been proven well-suited to describing interactions, has not been applied for inference. CDRVAE-M performs better than CDRVAE-D and presents the second-best performance, indicating the effectiveness of multinomial likelihood. Since the factors influencing user interests are diverse, Mult-VAE with multinomial mixture distribution is more appropriate for recommendation tasks than DVAE. However, this approach has a simpler encoder structure than CDRVAE. It does not fully consider the posterior collapse problem, which may limit the model to learning inherent properties of data and restrict the further optimization of the model. This may be why this variant has relatively poor performance compared to CDRVAE. Through the above analysis, the hybrid architecture does work in the proposed model. Combines denoising function and multinomial likelihood and alleviates the posterior collapse with a complex encoder. The hybrid architecture shows superior behaviour capturing than DVAE and Mult-VAE.

TABLE 7: Performance comparison $w.r.t$ different VAE architecture on Books&Music.

Methods	Books			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
CDRVAE	**0.5891**	**0.3975**	**0.3423**	**0.6095**	**0.4161**	**0.3527**
CDRVAE-V	0.5014	0.3119	0.2601	0.5008	0.3178	0.2590
CDRVAE-D	0.5666	0.3934	0.3318	0.5751	0.3969	0.3420
CDRVAE-M	0.5873	0.3945	0.3375	0.6045	0.4080	0.3506

5.8 Impact of Transfer Factor (RQ4)

To investigate the effect of transferring information, a large number of experiments have been conducted on hyper-parameter η. We adjust η between 0 and 1 in 0.1 intervals to observe the difference in performance. When η equals 0, the cross-domain recommendation model is equivalent to a single-domain model. When η is 1, information from both domains will be treated with equal importance. We test on three benchmark datasets and calculate HR metrics based on the first ten returned items. Fig. 4, Fig. 5 and Fig. 6, respectively describe the effect of the cross-domain factor η on recommendation results.

Fig. 4 and Fig. 5 show a similar tendency. With the increase of the parameter value, the recommended performance generally presents a trend of increasing first and then decreasing. When η is set as 0.2, the CDRVAE gives the best results. When the parameter is 0.5, the recommendation accuracy also has a slight improvement. Such experimental results reflect that additional knowledge from the auxiliary domain is necessary for cross-domain recommendation, and the transformation matrix does convert some beneficial cross-domain knowledge to another domain. Additionally, the knowledge weight of the auxiliary domain will affect the final results. Therefore, a balance between two domain information is critical for CDR when using supplemental information.

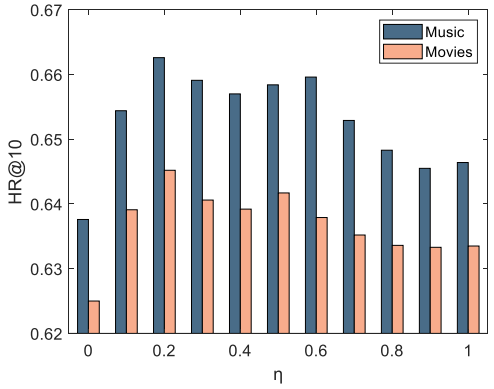

Fig. 4: Performance comparison $w.r.t$ different η on Movies&Books.

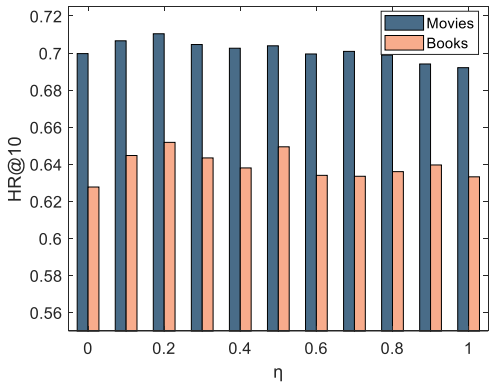

Fig. 5: Performance comparison $w.r.t$ different η on Movies&Music.

Fig. 6 depicts the results of the variation in model performance with different values of parameter η on Books&Music. As the proportion of the cross-domain loss increases, the recommendation results worsen, even lower than the results of the single-domain recommendation. This may be because the data density of the two domains in Books&Music is similar, so the model is more sensitive to η. With the increase of η, excessive auxiliary domain messages probably lead to negative transfer, which interferes with the utilization of information by the decoder.

CDRVAE applies the same decoder to learn both within-domain and cross-domain knowledge. If the transferred knowledge cannot be accurately converted into heterogeneous domain attributes, a higher cross-domain loss will generate more noise and affect the decoder's reconstruction of within-domain preferences. Through the above experiments, the value of η is fixed as 0.2 for better recommendations.

5.9 Impact of the Depth of Encoder and Decoder (RQ5)

As mentioned in section 4.2, we define the decoder as a simple single layer neural network to avoid posterior collapse. However, the proposed model expects the decoder to have reconstruction capability to predict unobserved user interaction entries. Thus, it's necessary to explore the impact

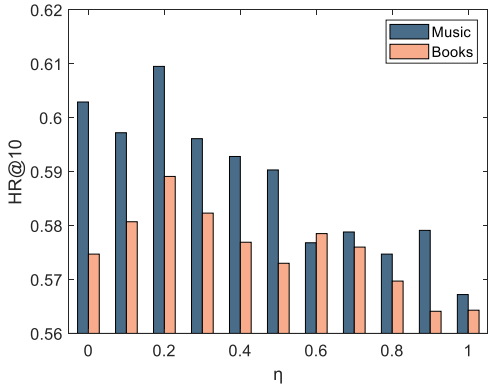

Fig. 6: Performance comparison $w.r.t$ different η on Books&Music.

of decoder depth on CDRVAE. We validate three datasets and calculate metrics based on the first ten returned items. The experimental results are listed in TABLE 8, TABLE 9 and TABLE 10.

TABLE 8: Performance comparison $w.r.t$ different decoder depth on Movies&Books.

Depth	Movies			Books		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
1	**0.7104**	**0.4629**	**0.4254**	**0.6518**	**0.4496**	**0.3871**
2	0.6761	0.4602	0.3940	0.6235	0.4205	0.3587
3	0.6765	0.4585	0.3911	0.6264	0.4277	0.3661

TABLE 9: Performance comparison $w.r.t$ different decoder depth on Movies&Music.

Depth	Movies			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
1	**0.6452**	**0.4259**	**0.3579**	**0.6626**	**0.4533**	**0.3889**
2	0.6172	0.3933	0.3250	0.6330	0.4205	0.3544
3	0.5942	0.3791	0.3145	0.6249	0.4098	0.3444

TABLE 10: Performance comparison $w.r.t$ different decoder depth on Movies&Books.

Depth	Books			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
1	**0.5891**	**0.3975**	**0.3423**	**0.6095**	**0.4161**	**0.3527**
2	0.5620	0.3650	0.3039	0.5729	0.3748	0.3142
3	0.5649	0.3735	0.3148	0.5713	0.3779	0.3219

We set the number of decoder layers to 1, 2 and 3 and test the prediction results separately. Clearly, we have the following observation: (1). The model performs best when the depth of the decoder is 1. (2). With the deepening of decoder layers, the model performance drops sharply. (3). When the depth of the decoder is 2 or 3, the difference in evaluation metrics is subtle. The above findings demonstrate that a single-layer decoder in CDRVAE can

reconstruct user preference distributions well. When the decoder gets multilayers, it introduces more parameters, resulting in an additional cost to achieve convergence. At the same time, more layers would give the decoder more representational power to reconstruct distribution from the noise sampled by the reparametrization process but would also lead to less reliance on the information derived from the inference network. The experimental results confirm that it is reasonable to simplify the decoder to one layer when the encoder of CDRVAE is determined.

The depth of the encoder affects the capture of user behaviour preference. We have also modified the number of encoder layers to complete the experiment. The statistic details are shown in the following TABLE 11, TABLE 12 and TABLE 13. From the experimental results, it can be found that as the number of encoder layers decreases, the recommendation performance also degrades. Since user preferences are affected by many factors, modeling behaviour distribution requires deeper encoder layers for complete mining.

TABLE 11: Performance comparison $w.r.t$ different encoder depth on Movies&Books.

Depth	Movies			Books		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
5	**0.7104**	**0.4629**	**0.4254**	**0.6518**	**0.4496**	**0.3871**
4	0.6970	0.4576	0.4205	0.6427	0.4465	0.3856
3	0.7031	0.4528	0.4189	0.6391	0.4414	0.3835

TABLE 12: Performance comparison $w.r.t$ different encoder depth on Movies&Music.

Depth	Movies			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
5	**0.6452**	**0.4259**	**0.3579**	**0.6626**	**0.4533**	**0.3889**
4	0.6332	0.4207	0.3537	0.6530	0.4504	0.3873
3	0.6391	0.4203	0.3520	0.6448	0.4434	0.3840

TABLE 13: Performance comparison $w.r.t$ different encoder depth on Books&Music.

Depth	Books			Music		
	HR@10	NDCG@10	MRR@10	HR@10	NDCG@10	MRR@10
5	**0.5891**	**0.3975**	**0.3423**	**0.6095**	**0.4161**	**0.3527**
4	0.5781	0.3927	0.3385	0.6002	0.4128	0.3513
3	0.5887	0.3972	0.3375	0.6024	0.4103	0.3507

5.10 Impact of Different Data Density and Overlapping User Scales(RQ6)

Interaction records are an important basis for predicting user behaviour. In practical scenarios, the sparsity in different product domains may vary greatly. It is necessary to explore whether CDRVAE performs better than other models. To validate the effect of the different density levels, we randomly sample records from the sparser product domain in the paired dataset at a rate of 80%, 60% and 40%, respectively and treat them as the new information source. As a result, the density of the datasets becomes 80%, 60% and 40% of the original. The paired dataset Movies&Music is selected for validation. The results are depicted in Fig. 7, Fig. 8 and Fig. 9.

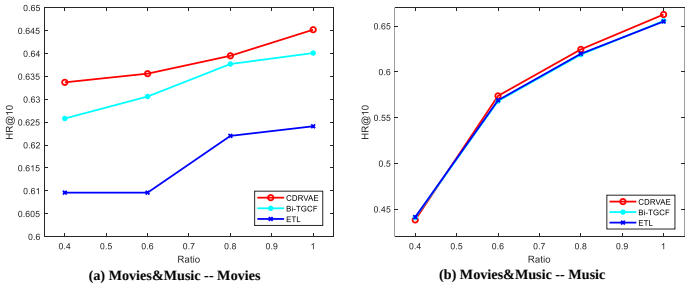

Fig. 7: Performance comparison $w.r.t$ different density levels with HR@10 on Movies&Music.

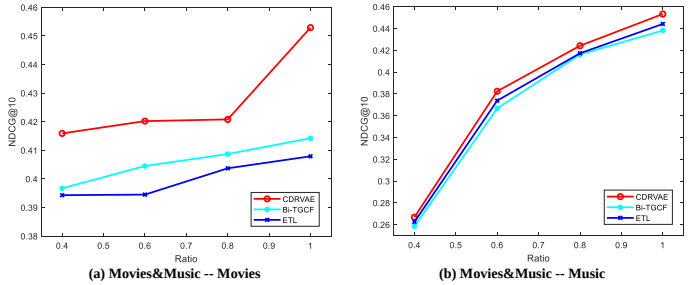

Fig. 8: Performance comparison $w.r.t$ different density levels with NDCG@10 on Movies&Music.

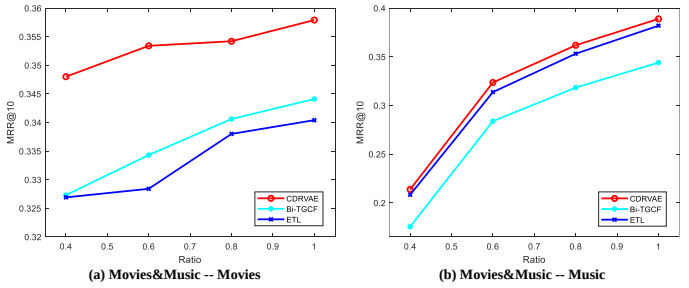

Fig. 9: Performance comparison $w.r.t$ different density levels with MRR@10 on Movies&Music.

From the experimental results, the following conclusion can be observed: (1). The performance in both domains becomes worse as the density level decreases. Because the reduction in data occurred in the music domain, the deterioration in metrics is more dramatic. (2). CDRVAE maintains a relatively superior performance. Even with the expansion of the density difference, the information that can be leveraged turns less. CDRVAE also remains competitive.

Since the existence of shared users between the two domains is an assumption of CDRVAE, it is necessary to

explore the impact of different numbers of shared users on CDRVAE. To better investigate this question, we choose Movies&Music and divide the number of overlapping users into four groups. The number of users in each group is 0.25, 0.5, 0.75 and 1 of the total number of overlapped users. The specific experimental results are shown in the following figure:

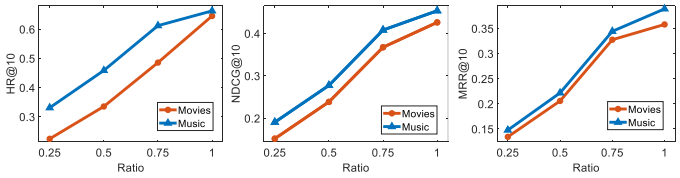

Fig. 10: Performance comparison $w.r.t$ different overlap user scales on Movies&Music.

We can conclude from the above figure that as the number of shared users decreases, the performance on both domains shows a decreasing trend. This conclusion can be explained by the fact that the number of shared users reflects the amount of information that can be transformed between two domains. In CDR, more users support better recommendation performance.

6 CONCLUSION

In this paper, we design a VAE-based cross-domain recommendation framework named CDRVAE. This approach significantly improves the performance of CDR for dual target domains. We implement this by applying a hybrid VAE architecture as the backbone for dual-path (both within-domain and cross-domain) preference modeling. CDRVAE learns powerful and plausible domain-specific latent factors via an asymmetric codec structure and the combined prior strategy, then employs a transformation matrix to extract potential mapping relationships. Finally, the domain-specific decoder reconstructs preference distributions from within domain latent variables and the converted latent variables simultaneously. Experiments demonstrate that the performance of this model is significantly superior to that of the others on three datasets.

Although CDRVAE shows potential in CDR, some experimental settings are based on empirical assumptions without sufficient theoretical analysis, such as the combined prior. We will explore more complicated prior and adaptive dynamic parameter selection to motivate further improvement. Additionally, CDR alleviates the data sparsity but still suffers from cold-start problems. We will also attempt to combine CDRVAE and other techniques for user cold-start and item cold-start recommendations.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant (Nos. 62071354 and 62201419), the Open Research Fund from Guangdong Laboratory of Artiffcial Intelligence and Digital Economy (Grant No. GML-KF-22-01), the National Natural Science Foundation of Shaanxi Province (Nos. 2019ZDLGY03-03) and also supported by the ISN State Key Laboratory.

REFERENCES

[1] Jyoti Aneja, Alex Schwing, Jan Kautz, and Arash Vahdat. A contrastive learning approach for training variational autoencoder priors. *Advances in Neural Information Processing Systems*, 34, 2021.
[2] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. Cross-domain mediation in collaborative filtering. In *International Conference on User Modeling*, pages 355–359. Springer, 2007.
[3] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In *International conference on machine learning*, pages 1613–1622. PMLR, 2015.
[4] Samuel R Bowman, Luke Vilnis, Oriol Vinyals, Andrew M Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. In *20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016*, pages 10–21. Association for Computational Linguistics (ACL), 2016.
[5] Xu Chen, Ya Zhang, Ivor Tsang, Yuangang Pan, and Jingchao Su. Towards equivalent transformation of user preferences in cross domain recommendation. *arXiv preprint arXiv:2009.06884*, 2020.
[6] /colorAhangama, Sapumal and Poo, Danny Chiang-Choon. /colorLatent user linking for collaborative cross domain recommendation. */colorarXiv preprint arXiv:1908.06583*, /color2019.
[7] /colorYu, Xu and Hu, Qiang and Li, Hui and Du, Junwei and Gao, Jia and Sun, Lijun. /colorCross-domain recommendation based on latent factor alignment. */colorNeural Computing and Applications*, /color34(/color5):/color3421–3432, /color2022.
[8] Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, and Francesco Ricci. Cross-domain recommender systems: A survey of the state of the art. In *Spanish conference on information retrieval*, pages 1–12. sn, 2012.
[9] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.
[10] Guangneng Hu, Yu Zhang, and Qiang Yang. Conet: Collaborative cross networks for cross-domain recommendation. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 667–676, 2018.
[11] Liang Hu, Jian Cao, Guandong Xu, Longbing Cao, Zhiping Gu, and Can Zhu. Personalized recommendation via cross-domain triadic factorization. In *Proceedings of the 22nd international conference on World Wide Web*, pages 595–606, 2013.
[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.
[13] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
[14] Pan Li and Alexander Tuzhilin. Ddtcdr: Deep dual transfer cross domain recommendation. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 331–339, 2020.
[15] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. Cccfnet: a content-boosted collaborative filtering neural network for cross domain recommender systems. In *Proceedings of the 26th international conference on World Wide Web companion*, pages 817–818, 2017.
[16] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference*, pages 689–698, 2018.
[17] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. Cross domain recommendation via bi-directional transfer graph collaborative filtering networks. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 885–894, 2020.
[18] Shuchang Liu, Shuyuan Xu, Wenhui Yu, Zuohui Fu, Yongfeng Zhang, and Amelie Marian. Fedct: Federated collaborative transfer for recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 716–725, 2021.
[19] Babak Loni, Yue Shi, Martha Larson, and Alan Hanjalic. Cross-domain collaborative filtering with factorization machines. In *European conference on information retrieval*, pages 656–661. Springer, 2014.
[20] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
[21] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. Cross-

domain recommendation: An embedding and mapping approach. In *IJCAI*, volume 17, pages 2464–2470, 2017.
[22] Nima Mirbakhsh and Charles X Ling. Improving top-n recommendation for cold-start users via cross-domain information. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(4):1–19, 2015.
[23] Andriy Mnih and Russ R Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2008.
[24] Aghiles Salah, Thanh Binh Tran, and Hady Lauw. Towards source-aligned variational models for cross-domain recommendation. In *Fifteenth ACM Conference on Recommender Systems*, pages 176–186, 2021.
[25] Huajie Shao, Shuochao Yao, Dachun Sun, Aston Zhang, Shengzhong Liu, Dongxin Liu, Jun Wang, and Tarek Abdelzaher. Controlvae: Controllable variational autoencoder. In *International Conference on Machine Learning*, pages 8655–8664. PMLR, 2020.
[26] Ilya Shenbin, Anton Alekseev, Elena Tutubalina, Valentin Malykh, and Sergey I Nikolenko. Recvae: A new variational autoencoder for top-n recommendations with implicit feedback. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 528–536, 2020.
[27] Ajit P Singh and Geoffrey J Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 650–658, 2008.
[28] Shulong Tan, Jiajun Bu, Xuzhen Qin, Chun Chen, and Deng Cai. Cross domain recommendation based on multi-type media fusion. *Neurocomputing*, 127:124–134, 2014.
[29] Quoc-Tuan Truong, Aghiles Salah, and Hady W Lauw. Bilateral variational autoencoder for collaborative filtering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 292–300, 2021.
[30] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.
[31] Yao Wu, Christopher DuBois, Alice X Zheng, and Martin Ester. Collaborative denoising auto-encoders for top-n recommender systems. In *Proceedings of the ninth ACM international conference on web search and data mining*, pages 153–162, 2016.
[32] Haowen Xu, Wenxiao Chen, Jinlin Lai, Zhihan Li, Youjian Zhao, and Dan Pei. On the necessity and effectiveness of learning the prior of variational auto-encoder. *arXiv preprint arXiv:1905.13452*, 2019.
[33] Feng Yuan, Lina Yao, and Boualem Benatallah. Darec: deep domain adaptation for cross-domain recommendation via transferring rating patterns. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4227–4233, 2019.
[34] Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 654–664, 2017.
[35] Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. Dtcdr: A framework for dual-target cross-domain recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1533–1542, 2019.
[36] Feng Zhu, Yan Wang, Jun Zhou, Chaochao Chen, Longfei Li, and Guanfeng Liu. A unified framework for cross-domain and cross-system recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
[37] Qile Zhu, Wei Bi, Xiaojiang Liu, Xiyao Ma, Xiaolin Li, and Dapeng Wu. A batch normalized inference network keeps the kl vanishing away. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2636–2649, 2020.
[38] Yongchun Zhu, Kaikai Ge, Fuzhen Zhuang, Ruobing Xie, Dongbo Xi, Xu Zhang, Leyu Lin, and Qing He. 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*, pages 1813–1817, 2021.

Tong Zhang received the B.E. degree in electronic information engineering from China University of Geosciences (Beijing), Beijing, China, in 2020. She is currently working toward the MS degree in information and telecommunication engineering with the School of Telecommunications Engineering, Xidian University, Xi'an, China. Her research interests include deep learning, information retrieval and recommender systems.

Chen Chen received the B.E. degree in communication engineering from Lanzhou University of Technology, Lanzhou, China, in 2019. He is currently pursuing the Ph.D. degree in information and telecommunication engineering with the School of Telecommunications Engineering, Xidian University, Xi'an, China. His current research interests include deep learning, recommender systems, and cross modal retrieval.

Dan Wang received her Ph.D. degree in communication and information systems from Xidian University, Xi'an, China in 2021. From 2019 to 2020, she is a visiting Ph.D. student with the Department of Electrical and Computer Engineering, Carleton University, Ottawa, ON, Canada. She is currently a Lecturer with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an. Her current research interests include edge learning, reinforcement learning, information retrieval and recommender systems.

Jie Guo received the B.E. degree in communication engineering from Zhengzhou University, Zhengzhou, China, in 2011 and the Ph.D. degree from Xidian University, Xian, China, in 2017. She is currently an associate professor with Xidian University. From 2015 to 2016, she got the state scholarship fund from China Scholarship Council to be an exchange Ph.D. Student with Carleton University, Canada. Her research interests include information fusion, deep learning, and recommender systems.

Bin Song (Senior Member, IEEE) received his BS, MS, and PhD in communication and information systems from Xidian University, Xi'an, China in 1996, 1999, and 2002, respectively. He is currently a professor of information and telecommunication engineering at the Xidian University, Xi'an, China. He has authored over 80 journal papers or conference papers and 40 patents. His research interests and areas of publication include multimedia communication, multimodal data fusion, content-based image recognition and machine learning, reinforcement learning, Internet of Things, big data, recommender systems.