10-2021

Towards source-aligned variational models for cross-domain recommendation

Aghiles SALAH
Singapore Management University, asalah@smu.edu.sg

Thanh-Binh TRAN
Singapore Management University, tbtran@smu.edu.sg

Hady W. LAUW
Singapore Management University, hadywlauw@smu.edu.sg

Citation

SALAH, Aghiles; TRAN, Thanh-Binh; and LAUW, Hady W.. Towards source-aligned variational models for cross-domain recommendation. (2021). *RecSys'21: Proceedings of the 15th ACM Conference on Recommender Systems, September 27 - October 1, Virtual*. 176-186.
Available at: https://ink.library.smu.edu.sg/sis_research/6430

Towards Source-Aligned Variational Models for Cross-Domain Recommendation

Aghiles Salah*
aghiles.salah@rakuten.com
Rakuten Institute of Technology
Paris, France

Thanh-Binh Tran
ttbinh92@gmail.com
Singapore Management University
Singapore, Singapore

Hady W. Lauw
hadywlauw@smu.edu.sg
Singapore Management University
Singapore, Singapore

ABSTRACT

Data sparsity is a long-standing challenge in recommender systems. Among existing approaches to alleviate this problem, cross-domain recommendation consists in leveraging knowledge from a source domain or category (e.g., Movies) to improve item recommendation in a target domain (e.g., Books). In this work, we advocate a probabilistic approach to cross-domain recommendation and rely on variational autoencoders (VAEs) as our latent variable models. More precisely, we assume that we have access to a VAE trained on the source domain that we seek to leverage to improve preference modeling in the target domain. To this end, we propose a model which learns to fit the target observations and align its hidden space with the source latent space jointly. Since we model the latent spaces by the variational posteriors, we operate at this level, and in particular, we investigate two approaches, namely rigid and soft alignments. In the former scenario, the variational model in the target domain is set equal to the source variational model. That is, we only learn a generative model in the target domain. In the soft-alignment scenario, the target VAE has its variational model, but which is encouraged to look like its source counterpart. We analyze the proposed objectives theoretically and conduct extensive experiments to illustrate the benefit of our contribution. Empirical results on six real-world datasets show that the proposed models outperform several comparable cross-domain recommendation models.

KEYWORDS

Cross-Domain Recommendation, Variational Autoencoder, Collaborative Filtering, Neural Networks

ACM Reference Format:
Aghiles Salah, Thanh-Binh Tran, and Hady W. Lauw. 2021. Towards Source-Aligned Variational Models for Cross-Domain Recommendation. In *Fifteenth ACM Conference on Recommender Systems (RecSys '21), September 27-October 1, 2021, Amsterdam, Netherlands.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3460231.3474265

*A major part of the work was done while the author was with Singapore Management University.

RecSys '21, September 27-October 1, 2021, Amsterdam, Netherlands

ACM ISBN 978-1-4503-8458-2/21/09…$15.00
https://doi.org/10.1145/3460231.3474265

1 INTRODUCTION

Recommender systems play an essential role in helping the users of modern applications to navigate the tremendous amount of choices offered to them [8, 43]. A standard approach to recommendation consists in fitting a statistical model to observed user-item interactions (e.g., ratings, clicks) that can be generalized to estimate unknown user-item preferences [14, 19, 29, 36, 37, 48]. This way of solving the recommendation problem is known as model-based collaborative filtering in the literature [5, 10, 24], and enjoys two main benefits, namely high performance and scalability.

One perennial challenge for the family of collaborative filtering algorithms is data sparsity, i.e., the number of observed interactions for every user or item is typically much lower than the possible number of interactions. It is not uncommon for data sparsity to reach 98% or more, i.e., for an average user we observe her interactions with merely 2% of all items yet seek to predict the rest. In the extreme cold-start scenario, which we consider in this paper, we may even seek to recommend items to users with no prior observations at all in a particular domain. This poses significant model estimation and generalization difficulties, as collaborative filtering relies on user-item interactions as the primary driver.

Problem. For those users who start cold in a particular domain (hereinafter known as the *target domain*), one approach is to bring in information from a different domain (hereinafter known as the *source domain*) where those users have some observed interactions. This approach is known as *cross-domain* recommendation [11, 13, 21, 28]. In this work, we focus on the scenario in which domains are item categories (e.g., Movies, Books) that share the same set of users. Under this setting, given a group of users, the goal is to use their observations and model in the source domain (e.g., Movies) to provide them with recommendations in the target domain (e.g., Books).

To meet this objective, different methods have been investigated in the literature (see Section 2). For instance, collective matrix factorization jointly factorizes rating matrices from various domains with a shared user-latent space [42]. There are also techniques inspired by transfer learning, which seek to either learn a mapping between the source and target models or infer user representations that are domain-invariant [26, 49]. Even so, cross-domain recommendation remains a relatively under-explored topic owing to its challenging nature.

Contributions. In this work, we advocate a probabilistic approach to cross-domain recommendation contrary to existing methods that are mostly deterministic. As our building block to model observations, we rely on deep generative latent variable models, namely variational autoencoders (VAEs) [18, 30], which have recently and evidentially shown strong performance on the item recommendation task [20, 24, 25, 39, 46]. More precisely, we leverage a VAE trained on the source domain to improve preference modeling in the target domain. To this end, we propose a variational autoencoder model which learns to fit the target observations and align its hidden space with the source latent space simultaneously (see Section 3). Since we model the latent spaces by the variational posteriors, we operate at this level. In particular, we investigate two approaches, namely rigid and soft alignments. In the former scenario, the variational model in the target domain is set equal to the source variational model. That is, we only learn a generative model in the target domain. In the soft-alignment scenario, the target VAE has its variational model, but which is encouraged to look like its source counterpart. In this latter scenario, our VAE modeling assumptions in the target domain give rise to a new objective, which we analyze theoretically to illustrate its properties for cross-domain recommendation.

Moreover, we conduct extensive experiments on six real-world datasets derived from a couple of e-commerce providers (see Section 4). Each dataset involves a pair of source and target domains from the same e-commerce provider, thus sharing a common user base which is a key vehicle to crossing domains. Comparison to several existing cross-domain recommendation methods showcase the utility of our proposed modeling approach.

2 RELATED WORK

Model-based collaborative filtering is an active research topic with a very rich literature. In this section, we briefly review work related to our contribution.

Matrix factorization (MF) models are predominant in collaborative filtering [9, 19, 33, 36, 38]. Several variants exists and the core differences are the assumptions on the latent and observation spaces [6, 32, 36], as well as the type of training objective, namely pointwise (scoring) loss [14, 36] or pairwise (ranking) loss [29]. Despite their success, the main limitation of MF-based models is their linear nature. This has motivated several contributions to develop non-linear recommendation models based on neural networks [10, 37, 48]. While these methods improve upon MF models in some situations, they turn out to be prone to overfitting when trained on sparse user-item preferences. More recently, deep generative models, namely variational autoencoders (VAEs) [18, 30], have shown strong performance improvement over several neural and MF-based models [24, 25, 39, 46]. This has motivated our choice of VAEs as the building blocks to model user-item preferences. These vanilla model-based recommender systems can be applied to the cross-domain task by concatenating the different domains to form a single dataset. However it turns out that this strategy could exacerbate the sparsity issue, which suggests that it is worth investigating models specifically designed for the cross-domain task.

Various cross-domain recommendation models have been proposed to alleviate the sparsity issue. While some earlier contribution were clustering-based [21], many cross-domain methods use matrix factorization as the base model to learn from observations. Among the popular approaches, collective matrix factorization jointly decomposes the rating matrices of the different domains with a shared user latent factors [42]. EMCDR, which stands for Embedding and Mapping framework for Cross-Domain Recommendation [26], fits MF models to the different domains and subsequently learns a mapping from the source to the target user embeddings or factors. To increase the modeling capacity, several efforts focused on building neural-based cross-domain recommendation methods [11, 22, 40, 49]. The notable Deep Domain Adaptation for Cross-Domain Recommendation (DaRec) [49] seeks to learn user representations that are domain invariant by relying on adversarial training. Both EMCDR and DaRec will be included as baselines.

More recently, motivated by the success of deep generative models in collaborative filtering, some authors rely on VAE to tackle cross-domain recommendation [1, 41]. The method proposed in [41] is very similar to EMCDR, it uses a Bayesian-VAE to learning the mapping between source and target user factors, which are obtained by fitting MF models to the different domains. This method is different from ours in several aspects. For instance, we use VAE to model the user preference, while the method in [41] uses matrix factorization. The model proposed in [1], which we refer to in this paper as Linked-VAEs, is the most closely related method to our contribution. Linked-VAEs jointly fits two VAEs, one to source data and another to the target data. To link the two VAEs, the outputs of both the source and target variational encoders are fed to the target decoder. In this paper we investigate an alternative VAE-based approach to cross-domain recommendation, which consists in encouraging the target variational model to align with its source counterpart. In our experiments, we include Linked-VAEs as a baseline to assess the importance of our contribution. To further elucidate the novelty of our approach, we will compare and contrast the modeling components of our proposed method vis-á-vis these related VAE-based models shortly in Section 3.

Note that there are also cross-domain recommendation models which integrates auxiliary data such as item textual descriptions or reviews [7, 12, 17, 23]. In this work, we do not use auxiliary data and learn from rating data only. In principle the method we propose can also be extended to incorporate various types of auxiliary data [47], which we leave to feature work.

3 METHOD

We assume that we are given a set of users along with their preferences over two disjoint sets of items, namely the source and target domains. In the rest of the paper we will use the superscripts "s" and "t" to refer to the source and target domains respectively. The user-item preferences in the source domain are organized into a matrix $\mathbf{X}^s = (x_{ui})$ of size $U \times I^s$, where x^s_{ui} denotes the interaction (e.g., rating) between user u and item i. The row of this matrix $\mathbf{x}^s_u$ is a vector containing all the preferences of user u. Similarly, $\mathbf{X}^t$ is the user-item matrix in the target domain.

3.1 Source Domain Model

We model the source domain data using a latent generative model of user preferences. We assume a continuous latent space with a standard isotropic Gaussian prior, i.e., $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{z} \in \mathbb{R}^K$. Conditional on the latent variables, the observations $\mathbf{x}^s_u$ are assumed to follow a Multinomial distribution parameterized using a neural network. That is,

$$p(\mathbf{x}_u^s|\mathbf{z}_u) = \text{Mult}(\mathbf{x}_{u\cdot}^s, \pi(\mathbf{z}_u)),$$
$$\text{with } \mathbf{z}_u \sim p(\mathbf{z}),\ \pi(\mathbf{z}_u) \propto \exp\{g_\theta(\mathbf{z}_u)\} \text{ and } x_{u\cdot}^s = \sum_i x_{ui} \quad (1)$$

where $g_\theta(\mathbf{z})$ is a neural network with parameters $\boldsymbol{\theta}$, which takes as input $\mathbf{z}_u$ and produces as output the unnormalized parameters of the Multinomial distribution $p(\mathbf{x}_u^s|\mathbf{z})$. Please refer to [24] for details on the benefits of using the Multinomial to model preference data.

To fit our model to observations, we need to infer the posterior $p(\mathbf{z}_u|\mathbf{x}_u^s)$, which is intractable. We therefore resort to Variational Inference (VI) [3, 4, 16], a popular and efficient approach to deal with complex probabilistic models. The idea of VI is to approximate the true posterior with a tractable variational model $q_\phi(\mathbf{z}_u|\mathbf{x}_u^s)$, governed by its own parameter ϕ. We let the variational distribution be a multivariate Gaussian with a diagonal covariance matrix:

$$q(\mathbf{z}_u|\mathbf{x}_u^s) = \mathcal{N}(\boldsymbol{\mu}_\phi(\mathbf{x}_u^s), \boldsymbol{\sigma}_\phi(\mathbf{x}_u^s)), \quad (2)$$

where $\boldsymbol{\mu}_\phi(\cdot)$ and $\boldsymbol{\sigma}_\phi(\cdot)$ are vector-valued functions – we use neural networks in this work – parameterized by ϕ, outputting respectively the mean and covariance parameters of the variational distributions. The combination of the above generative and inference models, $p(\mathbf{x}_u^s|\mathbf{z}_u)$ and $q_\phi(\mathbf{z}_u|\mathbf{x}_u^s)$, gives rise to a variational autoencoder for preference data [18, 24]. Learning under this framework amounts to maximizing the Evidence Lower BOund (ELBO), w.r.t. the model $\boldsymbol{\theta}$ and variational $\boldsymbol{\phi}$ parameters, which takes the following form,

$$\mathcal{L} = \sum_u \mathbb{E}_{q(\mathbf{z}_u|\mathbf{x}_u^s)}[\log p(\mathbf{x}_u^s|\mathbf{z}_u)] - \text{KL}(q_\phi(\mathbf{z}_u|\mathbf{x}_u^s)||p(\mathbf{z}_u)) \quad (3)$$

where KL stands for the Kullback–Leibler divergence. We use the *reparameterization trick* and rely on stochastic optimization to maximize the above objective, which is the standard procedure to estimate VAE models from data [18].

3.2 Target Domain Model

Our objective is to leverage the source data and model while making recommendation in the target domain. To this end, the core idea of our approach is to fit the target observations with a model whose latent space should align with the hidden space in the source domain. We let the target model have the same parametric form as the source model, and investigate two variants, namely rigid and soft alignments.

Rigid Alignment. We can achieve a perfect alignment between the source and target latent spaces by setting the variational model in the target domain equal to the source variational model. We refer to this model as Rigidly Aligned VAE (RA-VAE), and its graphical representation is depicted in Figure 1 (d). For learning, all we need is to estimate the target generative model, by maximizing the following objective,

$$\mathcal{L} = \sum_u \mathbb{E}_{q_{\phi^*}(\mathbf{z}_u|\mathbf{x}_u^s)}[\log p(\mathbf{x}_u^t|\mathbf{z}_u)] - \underbrace{\text{KL}(q_{\phi^*}(\mathbf{z}_u|\mathbf{x}_u^s)||p(\mathbf{z}_u))}_{Constant} \quad (4)$$

where the conditional likelihood $p(\mathbf{x}_u^t|\mathbf{z}_u)$ is a Multinomial defined as in (1), the superscript * is used to refer to the parameter values estimated from the source data, which are held fixed when fitting the target model. Hence, the KL-term in (4) is a constant. While the source and target latent spaces are identical, one limitation of this approach is that the lower bound of equation (4) may not be tight since the variational model is fixed, which could lead to a bad estimation of the generative model. We therefore investigate an alternative approach, which does not enforce a rigid alignment.

Soft Alignment. In this variant we seek to encourage the target variational model to look like its source counterpart. To do so in a principled way, we use the approximate posterior distribution $q_{\phi^*}(\mathbf{z}_u|\mathbf{x}_u)$ from the source domain as a prior over the latent variables in the target domain. That is, the target preferences of user u are assumed to have been generated from the following generative model $q_{\phi^*}(\mathbf{z}_u|\mathbf{x}_u^s)p(\mathbf{x}_u^t|\mathbf{z}_u)$, where $p(\mathbf{x}_u^t|\mathbf{z}_u)$ is defined as in (1) by substituting $\mathbf{x}_u^t$ for $\mathbf{x}_u^s$. We call this variant Softly Aligned VAE (SA-VAE), please refer to Figure 1 (e) for a graphical representation. Similar to the source model, we have to resort to variational inference to estimate our model from data. We introduce the following variational model, which as we shall see shortly, combined with the above specification of the generative model will give rise to an objective satisfying our desiderata.

$$q_\psi(\mathbf{z}_u|\mathbf{x}_u^t) = \mathcal{N}(\boldsymbol{\mu}_\psi(\mathbf{x}_u^t), \boldsymbol{\sigma}_\psi(\mathbf{x}_u^t)), \quad (5)$$

where $\boldsymbol{\mu}_\psi(\cdot)$ and $\boldsymbol{\sigma}_\psi(\cdot)$ are vector-valued functions – we use neural networks in this work – parameterized by ψ, outputting respectively the mean and covariance parameters of the variational distribution. Note that we choose a variational model where the conditioning on the observations is explicit only on the target data $\mathbf{x}_u^t$. The conditioning on source data is implicit, and more insights on this are provided in the next section. With the generative and variation inference models in place, the ELBO for the target model takes the following form,

$$\mathcal{L} = \sum_u \mathbb{E}_{q_\psi(\mathbf{z}_u|\mathbf{x}_u^t)}[\log p(\mathbf{x}_u^t|\mathbf{z}_u)] - \text{KL}(q_\psi(\mathbf{z}_u|\mathbf{x}_u^t)||q_{\phi^*}(\mathbf{z}_u|\mathbf{x}_u^s)) \quad (6)$$

It is now clear how the objective of equation (6) encourages the target variational model to be similar to the source variational model thanks to the KL-term. Note that by maximizing the above lower bound w.r.t. the target variational parameters ψ we can make it tighter. Similar to the vanilla VAE, we rely on the *reparameterization trick* and stochastic optimization to estimate the model $\boldsymbol{\theta}$ and variational parameters ψ from data. In the next section we will analyse this objective theoretically to gain more insight into some of its desirable properties for the cross-domain recommendation task.

3.3 Theoretical Analysis and Connection with the Information Bottleneck Principle

In this section, we provide an information-theoretic analysis of SA-VAE's criterion to gain more insights into some of its properties. Consider the following equivalent expression of (6), which is more convenient to work with for the purposes of our analysis.

$$\frac{1}{U} \times \mathcal{L} = \mathbb{E}_{q(\mathbf{x}^t,\mathbf{x}^s)q_\psi(\mathbf{z}|\mathbf{x}^t)}[\log p(\mathbf{x}^t|\mathbf{z})] - \text{KL}(q_\psi(\mathbf{z}|\mathbf{x}^t)||q_\phi(\mathbf{z}|\mathbf{x}^s)) \quad (7)$$

where $q(\mathbf{x}^t, \mathbf{x}^s)$ is the joint empirical data distribution of the source and target domain preferences, i.e., $q(\mathbf{x}^t) = \frac{1}{U}\sum_{u=1}^{U}\delta(\mathbf{x}^t - \mathbf{x}_u^t)\delta(\mathbf{x}^s -$

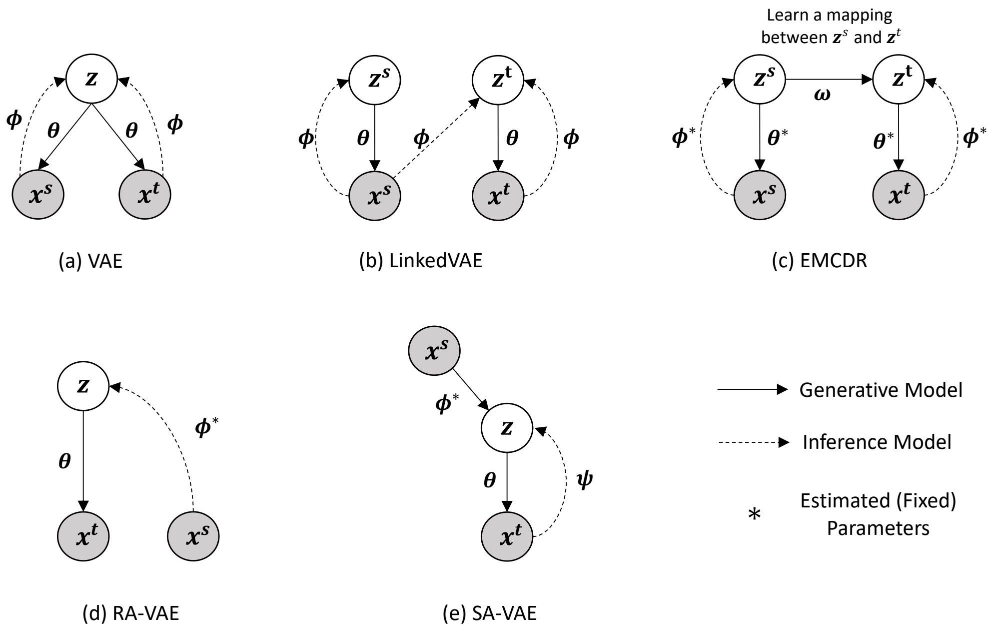

Figure 1: Graphical representation of different VAE-based cross-domain recommendation models. Note that EMCDR was initially used with matrix factorization, here we use it with VAE as the base model to learn user representations, which offers superior performance compared to using matrix factorization.

$\mathbf{x}_u^s$), with $\delta(\cdot)$ denoting the Dirac delta function, $q_\psi(\mathbf{z}|\mathbf{x}_u^t) \triangleq q_\psi(\mathbf{z}_u|\mathbf{x}_u^t)$ and $q_\phi(\mathbf{z}|\mathbf{x}_u^s) \triangleq q_\phi(\mathbf{z}_u|\mathbf{x}_u^s)$ by definition. With these new notations in place, we derive the following alternative expression of SA-VAE's objective function,

$$\frac{1}{U} \times \mathcal{L} = \mathbb{E}_{q(\mathbf{x}^t,\mathbf{x}^s)q_\psi(\mathbf{z}|\mathbf{x}^t)}[\log p(\mathbf{x}^t|\mathbf{z})] - \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) + \tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s) \quad (8)$$

where $\mathbb{I}_q(\mathbf{z}, \mathbf{x}^t)$ is the mutual information between $\mathbf{z}$ and $\mathbf{x}^t$ induced by the joint $q(\mathbf{z}, \mathbf{x}^t) = \int q(\mathbf{x}^t, \mathbf{x}^s) q_\psi(\mathbf{z}|\mathbf{x}^t) d\mathbf{x}^s$, while $\tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s)$ is a variational lower bound on the mutual information $\mathbb{I}_q(\mathbf{z}, \mathbf{x}^s)$ induced by the joint $q(\mathbf{z}, \mathbf{x}^s) = \int q(\mathbf{x}^t, \mathbf{x}^s) q_\psi(\mathbf{z}|\mathbf{x}^t) d\mathbf{x}^t$. We have expressed the KL-term in (7) as follows,

$$\mathrm{KL}(q_\psi(\mathbf{z}|\mathbf{x}^t)||q_\phi(\mathbf{z}|\mathbf{x}^s)) = \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) - \tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s). \quad (9)$$

The derivation details of expression (9) are given at the end of this section. Interestingly, the mutual information part in Equation (8) form a variational bound on the Information Bottleneck (IB) objective [2, 44, 45]. That is, for $\beta \leq 1$, we have

$$\mathrm{IB} = \mathbb{I}_q(\mathbf{z}, \mathbf{x}^s) - \beta \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) \geq \tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s) - \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) \quad (10)$$

Hence the proposed SA-VAE can be viewed as a generative model regularized with the Information Bottleneck principle. Recall that the Mutual information is a measure of mutual dependence between two variables. Equation (8) reveals that learning SA-VAE by maximizing its ELBO, encourages the latent variable $\mathbf{z}$ to be independent of the target observations $\mathbf{x}^t$ and to be dependant of the source data $\mathbf{x}^s$ simultaneously, while explaining the target observations thanks to maximizing the decoder model $p(\mathbf{x}^t|\mathbf{z})$. Hence, intuitively, as long as the model can explain the target observations, it prefers to rely more on the source data than the target data to infer the latent representations. This is a beneficial property in the context of cross-domain recommendation since, at test time, we seek to leverage the source domain model to make recommendation for new instances (users) in the target domain.

We now provide the derivation details of Equation (9).

Proof. Equation (8) can be obtained starting from (7) and rewriting the KL-term as follows.

$$\begin{aligned}
&\mathrm{KL}(q_\psi(\mathbf{z}|\mathbf{x}^t)||q_\phi(\mathbf{z}|\mathbf{x}^s)) \\
&= \mathbb{E}_{q(\mathbf{x}^t,\mathbf{x}^s)q_\psi(\mathbf{z}|\mathbf{x}^t)} \log \frac{q_\psi(\mathbf{z}|\mathbf{x}^t)}{q_\phi(\mathbf{z}|\mathbf{x}^s)} \\
&\overset{a}{=} \mathbb{E}_{q(\mathbf{x}^t,\mathbf{x}^s)q_\psi(\mathbf{z}|\mathbf{x}^t)} \log \frac{q_\psi(\mathbf{z}|\mathbf{x}^t) q(\mathbf{x}^t)}{q(\mathbf{x}^t) q_\psi(\mathbf{z})} \times \frac{q_\psi(\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x}^s)} \\
&\overset{b}{=} \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) + \mathbb{E}_{q(\mathbf{x}^t,\mathbf{x}^s)q_\psi(\mathbf{z}|\mathbf{x}^t)} \log \frac{q_\psi(\mathbf{z}) q(\mathbf{x}^s)}{q_\phi(\mathbf{z}|\mathbf{x}^s) q(\mathbf{x}^s)} \\
&= \mathbb{I}_q(\mathbf{z}, \mathbf{x}^t) - \tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s), \qquad (11)
\end{aligned}$$

where in a we have introduced the marginals $p(\mathbf{x}^t) = \int p(\mathbf{x}^t, \mathbf{x}^s) d\mathbf{x}^s$, $p_\psi(\mathbf{z}) = \int\int q(\mathbf{x}^t, \mathbf{x}^s) q_\psi(\mathbf{z}|\mathbf{x}^t) d\mathbf{x}^t d\mathbf{x}^s$, and in b we have introduced $p(\mathbf{x}^s) = \int p(\mathbf{x}^t, \mathbf{x}^s) d\mathbf{x}^t$. The term is $\tilde{\mathbb{I}}_q(\mathbf{z}, \mathbf{x}^s)$ is a variational lower bound on the mutual information $\mathbb{I}_q(\mathbf{z}, \mathbf{x}^s)$ since,

$$\begin{aligned}
&\mathbb{E}_{q(\mathbf{x}^s)} \int q_\psi(\mathbf{z}|\mathbf{x}^s) \log q_\psi(\mathbf{z}|\mathbf{x}^s) d\mathbf{z} \\
&\geq \mathbb{E}_{q(\mathbf{x}^s)} \int q_\psi(\mathbf{z}|\mathbf{x}^s) \log q_\phi(\mathbf{z}|\mathbf{x}^s) d\mathbf{z}. \qquad (12)
\end{aligned}$$

Inequality (12) can be easily demonstrated using the fact that

$$\mathrm{KL}(q_\psi(\mathbf{z}|\mathbf{x}^s)||q_\phi(\mathbf{z}|\mathbf{x}^s)) \geq 0 \quad (13)$$

Table 1: Data Statistics

Dataset			**#Users**	**#Items**		**#Ratings**		**Sparsity (%)**	
	Source	*Target*	*Shared*	*Source*	*Target*	*Source*	*Target*	*Source*	*Target*
Amazon	MoviesTV	Books	3,658	7,542	9,899	221,765	183,388	99.20	99.49
	MoviesTV	Musics	2,051	5,154	4,477	129,164	90,352	98.78	99.02
	Books	VideoGames	2 365	7,810	8,581	69,731	38,322	99,62	99,81
	Books	Musics	1,088	3,598	3,193	28,974	17,758	99,26	99,49
Douban	Movies	Books	10,638	18,403	15,691	2,283,322	633,797	98.83	99.62
	Movies	Musics	9,704	18,117	20,497	2,140,749	936,214	98.78	99.53

4 EXPERIMENTS

In this section we investigate the performance the proposed model on several real-world datasets. We conduct a systematic comparison with comparable cross-domain recommendation models.

4.1 Datasets

We use six publicly available datasets from Amazon [27] and Douban [15], which are commonly used in the context of cross-domain recommendation. Table 1 depicts these datasets and their statistics after preprocessing. Every dataset consists of a set of users along with their ratings across a pair of product categories representing the source and target domains. Following the common practice in related literature, we preprocess each dataset to keep only users with at least ten ratings. We further binarize the integer ratings by treating all available user-item interactions as positive feedback.

4.2 Comparative Baselines

We compare the proposed model with several cross-domain recommender models:

- **EMCDR** (Embedding and Mapping framework for Cross-Domain Recommendation) [26] learns a mapping between the source user latent space and target user latent space. This mapping is then leveraged to make recommendations for cold start users in the target domain for which we observe some ratings in the source domain. Note that, originally in EMCDR, the source and target user latent factors (embeddings) are obtained with matrix factorization. For parity in comparisons, in our experiment we use VAE to learn the user embeddings in the different domains. Figure 1 (c) depicts a graphical representation of our EMCDR baseline.
- **DARec** (Deep Domain Adaptation for Cross-Domain Recommendation) [49] is a recently proposed deep learning method, which tackles the problem of cross-domain recommendation by learning user representations that are domain-invariant.
- **LinkedVAEs** [1] fits two VAEs, one to the source data and another to the target data. For cross-domain recommendation purposes, the latent spaces of the two VAEs are linked by further feeding the user's source representation to the target decoder. The dependency structure of LinkedVAEs is illustrated in 1 (b).
- **VAE** (Variational Autoencoders for Collaborative Filtering) [24] has recently shown strong performance on the item recommendation task. To apply this model to the cross-domain recommendation task, we first concatenate the source and target domain preference matrices to form one dataset, i.e., $\mathbf{x} = [\mathbf{x}^t, \mathbf{x}^s]$. We then fit VAE to the resulting dataset $\mathbf{x}$. Figure 1 (a) provides a graphical illustration of this method, which seeks to learn a joint distribution of the source and target observations. Comparisons with VAE would allow us to gain insights to whether it is worth designing specific models for cross-domain recommendation. We rely on the VAE's implementation available in the Cornac recommendation framework [35].

4.3 Evaluation Metrics

We assess the recommendation accuracy on a set of held-out items (the test set) with two widely-used measures for top-M recommendation [5, 34].

NDCG@M (Normalized Discount Cumulative Gain) is an effective measurement for the quality of ranking. The DCG@M measures the gain of each item relative to its position in a list of top-M recommended items. Formally, for each user u,

$$DCG_u@M = \sum_{i \in D_u} \frac{\mathbb{1}[i \in L_u]}{\log(rank_i + 1)},$$

where D_u denotes the set of held-out items for user u, $\mathbb{1}[.]$ is the indicator function, $rank_i$ is the rank of item i in the top-M recommended items L_u, and

$$NDCG_u@M = \frac{DCG_u@M}{idealDCG_u@M},$$

where the $idealDCG_u@M$ is the best achievable $DCG_u@M$ in which all the held-out items are ranked at the top.

Recall@M denotes the ratio of correct items in a user's list of top-M recommendations to the number of her held-out items. Formally, for each user u,

$$Recall_u@M = \frac{\sum_{i \in D_u} \mathbb{1}[i \in L_u]}{|D_u|}.$$

To evaluate an entire model we average each measure over all users.

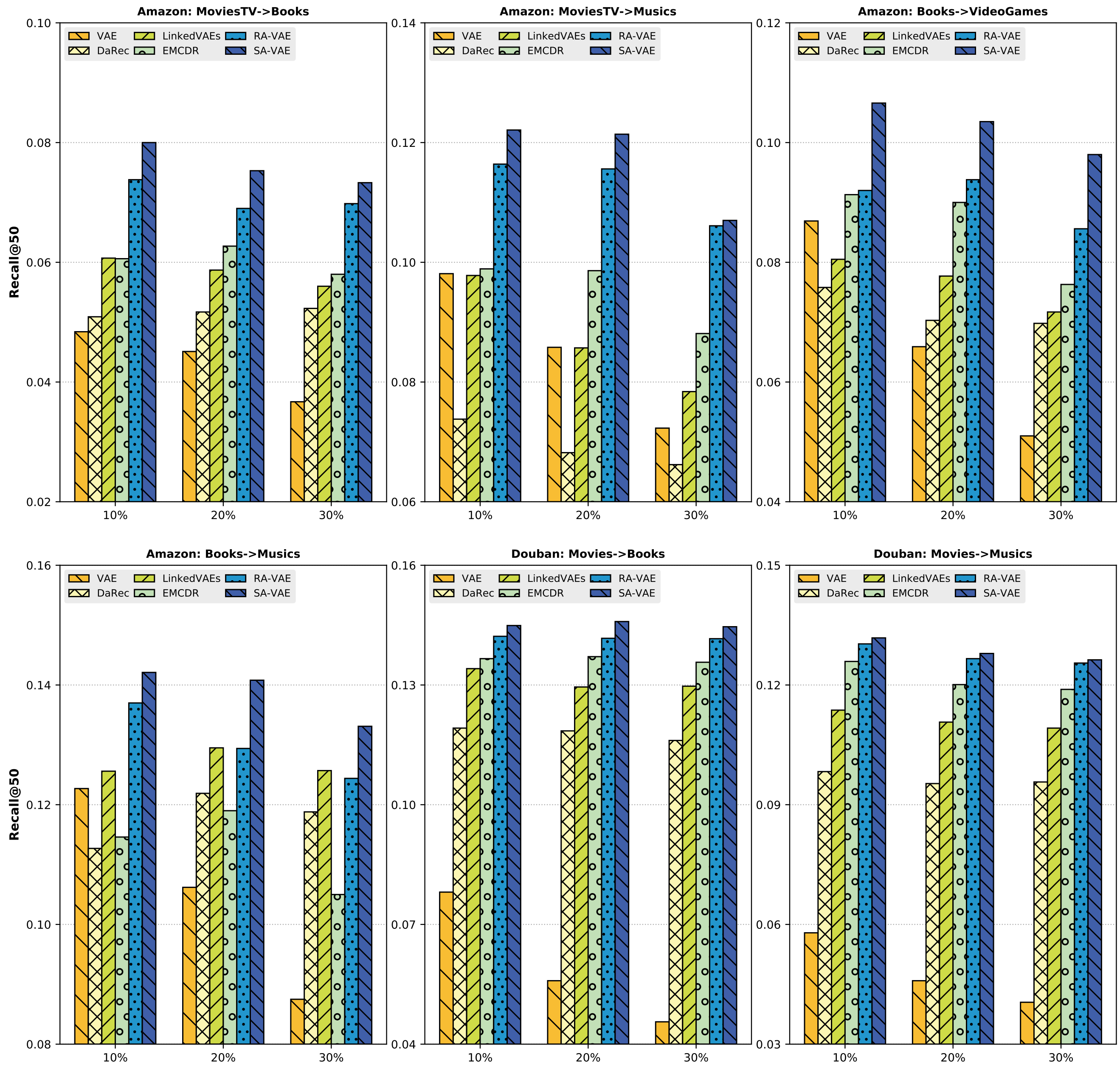

Figure 2: Recall@50 performance (y-axes) for every model, across different datasets and test set sizes (x-axes).

4.4 Experimental Settings

For training and testing under the cross-domain recommendation task, we follow the common approach, which consists in randomly splitting the target users into train and test sets. As in [26], the members of the test set are considered to be cold-start users in the target domain, i.e., all the target ratings of the test users are held-out for testing. This simulates the scenario where some users are active in some source domain and now begins anew in a target domain. We consider different proportions of such test users, namely 10%, 20% and 30% respectively. Hypothetically, the larger the proportion of the test users, the more challenging the task is, as there are fewer users in the training set to help inform the learning. We use (*Recall*) to tune the different models based on held-out validation sets (5% of training users).

The number of latent dimensions K for user representations is set to 20. We do not observe any significant change with higher numbers of latent dimensions. For the autoencoder family, we rely on multilayer perceptrons (MLPs) to parameterize the inference

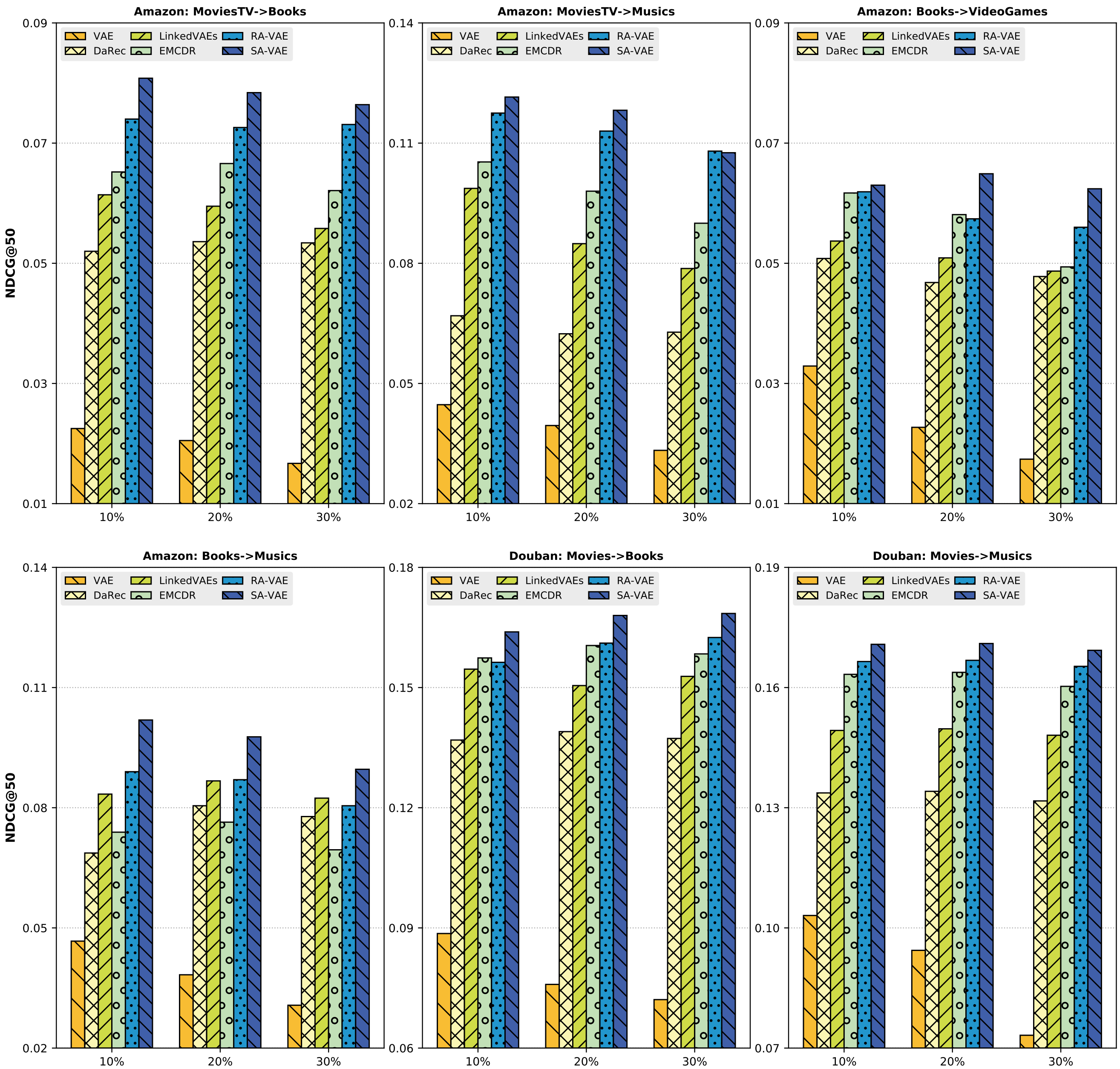

Figure 3: NDCG@50 performance (y-axes) for every model, across different datasets and test set sizes (x-axes).

and generative models. We explore MLPs with 0, 1, 2, and 3 hidden layers, finally retaining 1-hidden layer with 40 dimensions ($2 \times K$) for the inference models, and 0-hidden layer for the decoders. More layers provide only slight improvement, while introducing significant computational overhead. This is consistent with the findings of Liang et al. [24], who also recommend similar settings for VAE when applied to collaborative filtering. To introduce non-linearities at the hidden-layers, we use `Tanh` activation function, which we found to be superior to `ReLU` function. Since the observations are binary, we use `Sigmoid` activation function at the output of the different decoders.

To set the learning rate and number of training epochs, we perform a hyperparameter search. For the former, the search space is $\{1e^{-4}, \ldots, 1e^{-1}\}$ with multiples of 10. For the number of epochs the search space is $\{100, \ldots, 500\}$ with steps of 100. All models are fit to observation using stochastic gradient descent with the Adam update rule and a batch size of 128 instances. For any other remaining hyperparameter for the baseline methods, we follow the

Table 2: Summary of two-tailed paired t-test results (p-values) between the top three performing models.

		Amazon				*Douban*	
	Model Pair	**MovieTV Books**	**MovieTV Musics**	**Books VideoGames**	**Books Musics**	**Movies Books**	**Movies Musics**
Recall@50	SA-VAE RA-VAE	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$	< 0.05	0.08
	SA-VAE EMCDR	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$	< 0.05
	RA-VAE EMCDR	$< 0.01/6$	$< 0.01/6$	0.12	$< 0.01/6$	$< 0.01/6$	< 0.05
NDCG@50	SA-VAE RA-VAE	$< 0.01/6$	$< 0.01/6$	0.06	$< 0.01/6$	$< 0.01/6$	< 0.05
	SA-VAE EMCDR	$< 0.01/6$	$< 0.01/6$	0.07	$< 0.01/6$	$< 0.01/6$	$< 0.01/6$
	RA-VAE EMCDR	$< 0.01/6$	$< 0.01/6$	0.09	$< 0.01/6$	0.11	< 0.05

recommendations by the original authors of each corresponding model.

4.5 Results and Discussion

Figures 2 and 3 depict the performance of the competing models for different test-set sizes across all datasets in Recall and NDCG respectively. We repeat each experiment *thirty* times with different seeds and report the average result for each model. To assess the statistical significance of the results, we conduct two-tailed paired samples t-tests.

Overall, the proposed models, namely RA-VAE and SA-VAE offer noticeably higher performances than the baselines in most cases. The results hold across different test-sizes, and the improvements reached by the proposed models are statistically significant in the majority of cases. Table 2 summarizes the t-test results between the top three performing models, namely SA-VAE, RA-VAE and EMCDR, when 10% of target users are held-out for testing. For every pair of models, to account for multiple comparisons across different datasets [31], we consider the Bonferroni correction, which consists in testing each individual hypothesis at a significance level α/n, where α is the significance threshold and n is the number of hypotheses—corresponding to the number of datasets in our case. Besides the strong performance of our models, these results reveal some interesting insights, which we discuss below.

The VAE model exhibits the lowest results among baselines in many cases. Recall that to use this method under the cross-domain recommendation setting, we first concatenate the source and target domains to form a single dataset, and then we fit VAE to the resulting dataset. The concatenation operation however is likely to exacerbate data sparsity, which could explain the low performance of this baseline. This stresses the importance of tailoring recommendation model for the cross-domain setting.

Among the cross-domain baselines, while the patterns are more mixed, EMCDR seems to preform best in most cases, followed by LinkedVAEs, then DaRec. Interestingly, the former two models, EMCDR and LinkedVAEs, are also VAE-based as opposed to DaRec, which provide further support to using VAE as a building block to model observations. Our models consistently outperform the cross-domain baselines including the other VAE-based ones thereby demonstrating the significance of our approach.

Focusing on the proposed RA-VAE and SA-VAE variants, the performance is either statistically indistinguishable or SA-VAE is superior. One advantage of SA-VAE over RA-VAE is that the former variant has its variational model with free parameters. At each learning step, SA-VAE optimizes the ELBO w.r.t. the variational parameters to make this bound tighter to the true but intractable likelihood. RA-VAE however, uses the variational model from the source domain, which may result in a loose ELBO that can negatively impact the training of the target decoder model.

To investigate the robustness of our models regarding the number of observations per-user in the source domain, in Figure 4 we report the recommendation performance across several groups with different user-activity levels (number of ratings). The top two rows concern performance at Recall on the six datasets, while the bottom two rows concern the corresponding performance at NDCG. We retain the top three performing models only, namely SA-VAE, RA-VAE, and EMCDR, to avoid cluttering the figures. While the results are quite data-dependent, the proposed models consistently outperform EMCDR, which is the strongest baseline. It is also interesting to note that on some datasets (e.g., Amazon: Books->VideoGames and Douban: Movies->Musics) under low user-activity regimes (<10 ratings) our models, especially SA-VAE, improve substantially upon the performance of EMCDR. This suggests that the proposed models are more robust regarding user-activity in the source domain. Regarding the two proposed models, SA-VAE and RA-VAE, either

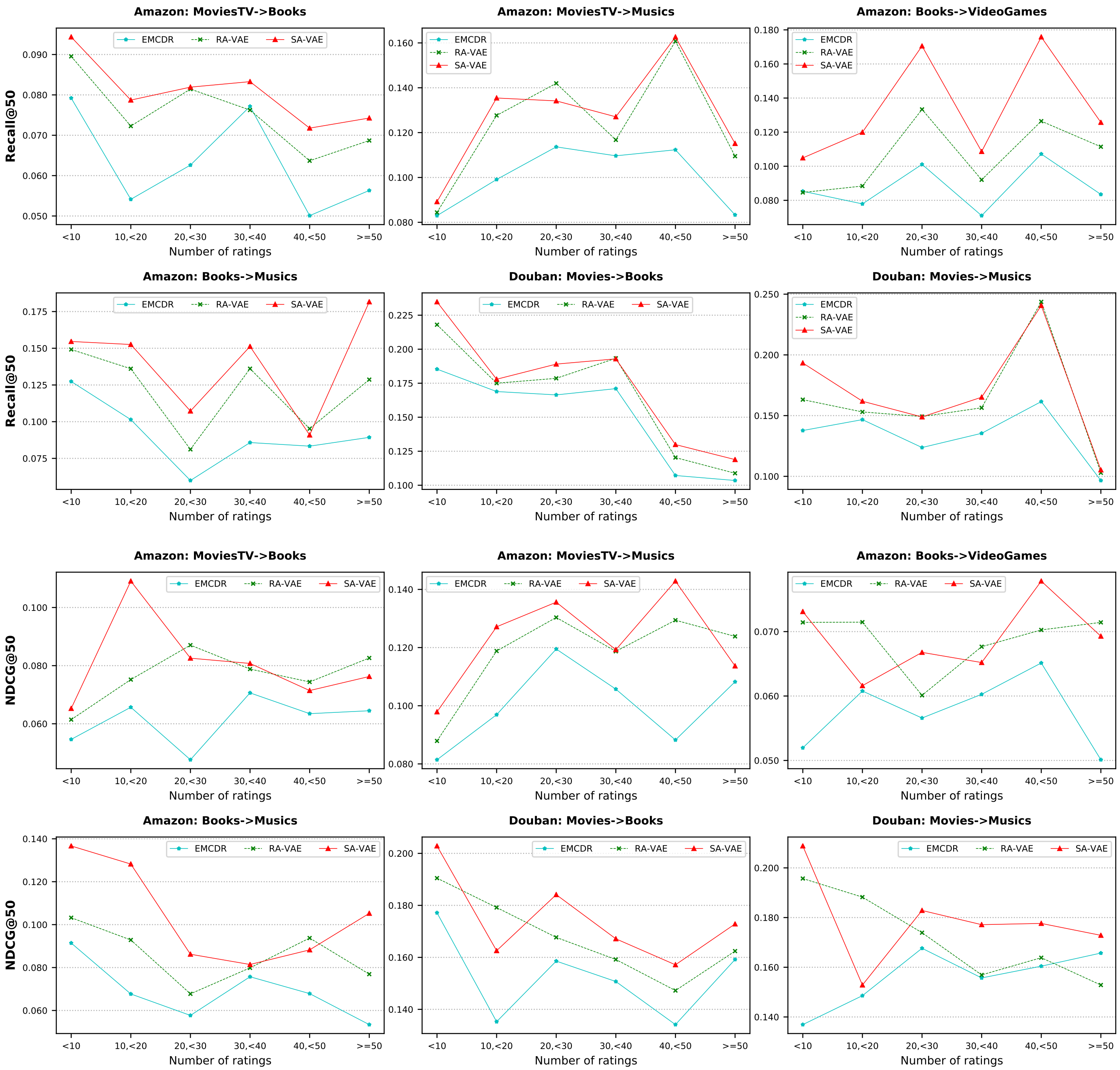

Figure 4: Recall and NDCG results across groups with different user-activity (number of observed preferences in the source domain).

the former outperforms the latter (in several cases) or the results are tight. The most important differences between the two models seem to occur in groups with a user-activity less than 30.

5 CONCLUSION

We propose a VAE-based approach for cross-domain recommendation. The core idea of our method is as follows. Given a VAE trained on the source domain, introduce another model in the target domain that simultaneously learns to fit observations and align its latent space with its source counterpart. We investigate two variants, namely rigid and soft alignments. In the former, rigid alignment scenario, the source and target latent space are identical. This is achieved by setting the variational model in the target domain equal to the source variational model. In the soft alignment case, the target variational model is encouraged to look like its source counterpart. We achieve this objective in a principled way by using the source variational model as a prior over the latent space of target domain. This gives rise to a new objective which enjoys useful properties for the cross-domain recommendation task as illustrated by our information-theoretic analysis. Empirical results on six real-world datasets demonstrate the importance of our contribution. The proposed models, noticeably outperform several strong cross-recommendation methods including neural and VAE-based ones. Moreover, our experiments also reveal some useful insights such as the importance of tailoring recommendation models for the cross-domain scenario. Among the two variants we propose, namely RA-VAE and SA-VAE, we find that the latter SA-VAE shows comparable or superior performance to RA-VAE.

Future work could include integrating auxiliary data (e.g., item content information) [47] to the proposed models to further alleviate the sparsity issue, tackle other cross-domain scenarios, such as the situation without user overlap between the source and target domains, or extend the ideas proposed in this paper to BiVAE [46] to address cross-domain recommendation on the user-side and item-side simultaneously.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its NRF Fellowship Programme (Award No. NRF-NRFF2016-07).

REFERENCES

[1] Sapumal Ahangama and Danny Chiang-Choon Poo. 2019. Latent User Linking for Collaborative Cross Domain Recommendation. *arXiv preprint arXiv:1908.06583* (2019).

[2] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. 2017. Deep variational information bottleneck. *International Conference on Learning Representations (ICLR)* (2017).

[3] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

[4] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. 2017. Variational Inference: A Review for Statisticians. *J. Amer. Statist. Assoc.* 112, 518 (Feb 2017), 859–877. https://doi.org/10.1080/01621459.2017.1285773

[5] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowl.-Based Syst.* 46 (2013), 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[6] Ali Taylan Cemgil. 2009. Bayesian Inference for Nonnegative Matrix Factorisation Models. *Comp. Int. and Neurosc.* 2009 (2009), 785152:1–785152:17. https://doi.org/10.1155/2009/785152

[7] 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.

[8] Carlos A Gomez-Uribe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2015), 1–19.

[9] Prem Gopalan, Jake M. Hofman, and David M. Blei. 2015. Scalable Recommendation with Hierarchical Poisson Factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI 2015, July 12-16, 2015, Amsterdam, The Netherlands*. 326–335. http://auai.org/uai2015/proceedings/papers/208.pdf

[10] 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, WWW 2017, Perth, Australia, April 3-7, 2017*. 173–182. https://doi.org/10.1145/3038912.3052569

[11] 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.

[12] Guangneng Hu, Yu Zhang, and Qiang Yang. 2019. Transfer meets hybrid: A synthetic approach for cross-domain collaborative filtering with text. In *The World Wide Web Conference*. 2822–2829.

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

[14] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008), December 15-19, 2008, Pisa, Italy*. 263–272. https://doi.org/10.1109/ICDM.2008.22

[15] Junming Huang, Xue-Qi Cheng, Hua-Wei Shen, Tao Zhou, and Xiaolong Jin. 2012. Exploring social influence via posterior effect of word-of-mouth recommendations. In *Proceedings of the fifth ACM international conference on Web search and data mining*. 573–582.

[16] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1999. An Introduction to Variational Methods for Graphical Models. *Machine Learning* 37, 2 (1999), 183–233. https://doi.org/10.1023/A:1007665907178

[17] Heishiro Kanagawa, Hayato Kobayashi, Nobuyuki Shimizu, Yukihiro Tagami, and Taiji Suzuki. 2019. Cross-domain recommendation via deep domain adaptation. In *European Conference on Information Retrieval*. Springer, 20–29.

[18] Diederik P. Kingma and Max Welling. 2014. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*. http://arxiv.org/abs/1312.6114

[19] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* 42, 8 (2009), 30–37. https://doi.org/10.1109/MC.2009.263

[20] Wonsung Lee, Kyungwoo Song, and Il-Chul Moon. 2017. Augmented Variational Autoencoders for Collaborative Filtering with Auxiliary Information. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*. 1139–1148.

[21] Bin Li, Qiang Yang, and Xiangyang Xue. 2009. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th annual international conference on machine learning*. 617–624.

[22] 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.

[23] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2017. 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*. 817–818.

[24] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*. 689–698. https://doi.org/10.1145/3178876.3186150

[25] Sam Lobel, Chunyuan Li, Jianfeng Gao, and Lawrence Carin. 2020. RaCT: Toward Amortized Ranking-Critical Training For Collaborative Filtering. In *International Conference on Learning Representations*.

[26] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach.. In *IJCAI*. 2464–2470.

[27] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference*. 43–52.

[28] Nima Mirbakhsh and Charles X Ling. 2015. Improving top-n recommendation for cold-start users via cross-domain information. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 9, 4 (2015), 1–19.

[29] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*. 452–461.

[30] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. 2014. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*. 1278–1286. http://proceedings.mlr.press/v32/rezende14.html

[31] G Rupert Jr et al. 2012. *Simultaneous statistical inference*. Springer Science & Business Media.

[32] Aghiles Salah and Hady W. Lauw. 2018. A Bayesian Latent Variable Model of User Preferences with Item Context. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI*, Jérôme Lang (Ed.). ijcai.org, 2667–2674.

[33] Aghiles Salah and Hady W Lauw. 2018. Probabilistic collaborative representation learning for personalized item recommendation. In *Uncertainty in Artificial Intelligence (UAI)*.

[34] Aghiles Salah and Mohamed Nadif. 2017. Social regularized von Mises–Fisher mixture model for item recommendation. *Data Mining and Knowledge Discovery* 31, 5 (2017), 1218–1241.

[35] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *J. Mach. Learn. Res.* 21 (2020), 95–1.

[36] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver,*

British Columbia, Canada, December 3-6, 2007. 1257–1264.

[37] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. AutoRec: Autoencoders Meet Collaborative Filtering. In *Proceedings of the 24th International Conference on World Wide Web Companion, WWW 2015, Florence, Italy, May 18-22, 2015 - Companion Volume*. 111–112. https://doi.org/10.1145/2740908.2742726

[38] Hanhuai Shan and Arindam Banerjee. 2010. Generalized Probabilistic Matrix Factorizations for Collaborative Filtering. In *ICDM 2010, The 10th IEEE International Conference on Data Mining, Sydney, Australia, 14-17 December 2010*. 1025–1030. https://doi.org/10.1109/ICDM.2010.116

[39] Ilya Shenbin, Anton Alekseev, Elena Tutubalina, Valentin Malykh, and Sergey I Nikolenko. 2020. 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*. 528–536.

[40] Shirish Shevade, MN Murty, et al. 2019. Neural Cross-Domain Collaborative Filtering with Shared Entities. *arXiv preprint arXiv:1907.08440* (2019).

[41] Jiarui Shi and Quanmin Wang. 2019. Cross-domain variational autoencoder for recommender systems. In *2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT)*. IEEE, 67–72.

[42] 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.

[43] Brent Smith and Greg Linden. 2017. Two decades of recommender systems at Amazon. com. *Ieee internet computing* 21, 3 (2017), 12–18.

[44] Naftali Tishby, Fernando C Pereira, and William Bialek. 1999. The information bottleneck method. (1999), 368–377.

[45] Naftali Tishby and Noga Zaslavsky. 2015. Deep learning and the information bottleneck principle. In *2015 IEEE Information Theory Workshop (ITW)*. IEEE, 1–5.

[46] Quoc-Tuan Truong, Aghiles Salah, and Hady W Lauw. 2021. Bilateral Variational Autoencoder for Collaborative Filtering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 292–300.

[47] Quoc-Tuan Truong, Aghiles Salah, Thanh-Binh Tran, Jingyao Guo, and Hady W Lauw. 2021. Exploring Cross-Modality Utilization in Recommender Systems. *IEEE Internet Computing* (2021).

[48] Yao Wu, Christopher DuBois, Alice X. Zheng, and Martin Ester. 2016. Collaborative Denoising Auto-Encoders for Top-N Recommender Systems. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, February 22-25, 2016*. 153–162.

[49] Feng Yuan, Lina Yao, and Boualem Benatallah. 2019. DARec: Deep Domain Adaptation for Cross-Domain Recommendation via Transferring Rating Patterns. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 4227–4233.