

跨域推荐中用户偏好等价转换论文的复现报告

XU CHEN, YA ZHANG, IVOR W. TSANG, YUANGANG PAN, JINGCHAO SU

摘要

跨域推荐(CDR)是推荐系统中的一个热门研究课题。该文主要讨论CDR的一个流行场景,其中不同的域共享同一组用户,但没有重叠的物品。在该文中,假设每个用户在一个领域的偏好可以用另一个领域来表示,并且这些偏好可以通过所谓的等效转换相互转换,学习用户偏好的两个特性。基于这一假设,提出一个等效转换学习器(ETL),它对用户行为跨域的联合分布进行建模。ETL中的等效转换对共享用户表征的思想进行了泛化,允许在不同域中学习到的偏好保留特定于领域的特征以及重叠的特征。本复现报告在三个公共数据集上进行的原始实验、不同潜在向量维度的实验以及不同的转换方式实验,结果与原论文报告结果一致。

关键词: 跨域推荐; 等效转换学习器; 变分推理

1 引言

跨域推荐(Cross domain recommendation, CDR) 通过从其他域迁移知识来解决推荐中的数据稀疏性问题,提高推荐的性能。用户是共享的,来自不同域的物品没有重叠是一个流行的场景。它在亚马逊或淘宝等现实世界的推荐系统中有实际应用,在这些系统中,同一用户可以购买涉及不同域的产品(如电影和书籍)。

在该文中,试图以更有原则的方式学习CDR的重叠和特定于域的特性。特别是,假设每个用户在不同域的偏好可以通过等效转换相互转换。然后,提出了一个等效转换学习器(ETL),它对用户行为跨域的联合分布进行建模^[2]。ETL中的等效转换了共享用户表征的思想,并使跨域的用户偏好具有保留特定于域的特征以及学习重叠特征的能力。这些贡献总结如下:

- ★强调了对CDR跨域用户行为联合分布建模的重要性;

- ★提出了一个等价的转换假设,并进一步提出了一种新的方法,称为ETL,在联合分布匹配方案中对重叠和特定域的特征进行建模。提出的模型以一种更有原则的方式工作,不需要选择训练用户或直观地设计网络;

- ★本复现报告在三个公共数据集上进行的原始实验、不同潜在向量维度的实验以及不同的转换方式实验,结果与原论文报告结果一致。

2 相关工作

推荐系统越来越受关注,其中CDR作为缓解推荐稀疏性问题的一种技术而出现^[1]。从跨域推荐的发展进行说明。此外,为了更好地说明提出的方法,这里还引入了深度生成建模。

2.1 跨域推荐

跨域推荐探索来自多个领域的的数据，以同时提高模型在多个场景下的推荐性能^[8]。一个基本的想法是结合跨领域知识的几个约束来同时分解两个领域中的用户-项目交互矩阵^[6,14]。这种方法可以在大量基于矩阵分解的单域推荐上进行扩展，但其性能不如深度学习方法。另一种范式是定制一个映射函数，其优化目标是转换后的冷启动用户表示在目标域内泛化良好^[15]。该范式的有效性取决于映射函数的合理性和表征能力，以及是否有足够的重叠实体可供训练，这限制了模型的泛化能力。第三种范式采用流行的知识图技术^[17]，该技术构建共享图来表示用户、项目和属性之间的关系，并通过图嵌入学习实体表示^[12]。尽管具有出色的图结构提取能力，但对计算资源的要求较高使这些方法的伸缩性受到潜在限制。最近，利用重叠用户表示和组合的算法很流行，它们的标准做法是从各个领域学习实体表示，然后将重叠的实体表示组合起来，以丰富每个领域的知识^[19]。显然，缺乏跨域相似度和粗糙的组合方式限制了它们的性能。

综上所述，以往的方法主要基于共享用户表示的思想，重点学习重叠的特征。在该文中，ETL建议学习CDR的重叠特性和特定域的特性。与最近通过依赖于所选用户行为的MLP映射函数来学习特定域特征的工作相比^[13]，所提出的ETL不需要仔细选择训练样本。当考虑基于侧信息的CDR工作(例如，物品审查，用户简介)时，的方法侧重于通过使用用户-物品交互来更好地转移知识。

2.2 深度生成建模

深度生成建模是一种强大的数据分布建模技术。它已被各种方法广泛讨论，并扩展到许多应用中。近年来，变分推理^[7]和生成对抗学习^[4]是深度生成建模的两种代表性技术。变分推理以概率的方式与指定的目标函数进行分布匹配。例如，VAE^[10]引入Kullback Leibler(KL)散度来匹配潜在后验分布与高斯先验。基于VAE，提出了各种变体^[18]来提高编码和重构性能，以及更好的分布建模方式。如Zhao等人^[18]提出了目标函数中相互信息最大化的InfoVAE，提高了变分后验的质量，并有助于更好的分布建模。GAN^[11]是另一种热门的深度生成建模技术。GAN包含一个生成器和一个鉴别器，其中鉴别器试图区分真实样本和假样本，而生成器试图混淆鉴别器。GAN中的对抗学习的优点是通过二元分类以更优雅的方式测量分布距离，并将研究人员从定义棘手的目标函数的痛苦实践中解放出来。

这些VAE+GAN的工作都是基于一个共享潜在空间的假设来进行分布匹配。然而，在CDR中，跨域的用户行为是相关的，但并不完全重叠，特定于域的特征对推荐很重要。在ETL中，提出了一种等效的基于变换的分布匹配方案。

3 方法

3.1 问题定义

表1总结了本文中的主要符号。

表 1. 主要符号

符号	描述
\mathcal{U}	跨域推荐用户集
\mathcal{X} and \mathcal{Y}	两个不同域
$I^{\mathcal{X}}$	\mathcal{X} 域的物品集
$I^{\mathcal{Y}}$	\mathcal{Y} 域的物品集
N	用户数
M	\mathcal{X} 域的物品数
T	\mathcal{Y} 域的物品数
$R^{\mathcal{X}}$	\mathcal{X} 域的用户行为矩阵
$R^{\mathcal{Y}}$	\mathcal{Y} 域的用户行为矩阵
x_i	\mathcal{X} 域中用户 i 的行为
y_i	\mathcal{Y} 域中用户 i 的行为
E_x, E_y	不同域的用户编码器
D_x, D_y	不同域的用户潜在编码的解码器
z_x	\mathcal{X} 域中用户的潜在编码
z_y	\mathcal{Y} 域中用户的潜在编码
W_x, W_y and W	近似等效转换的可训练矩阵
$\mathcal{D}_x, \mathcal{D}_y$	对抗性分布匹配中的鉴别器
d	潜在嵌入维度
λ	等效变换约束的超参
η	\mathcal{L}_{PRL} 超参
ϕ_x, ϕ_y	编码器参数
θ_x, θ_y	解码器的参数
ψ_x, ψ_y	鉴别器的参数
\mathcal{L}_{JRL}	联合重建损失
\mathcal{L}_{PRL}	先验正则化损失
\mathcal{L}_{ETL}	ETL的目标函数

在本文,不同的域 \mathcal{X} 和 \mathcal{Y} 具有相同的用户集 $\mathcal{U} = \{U_1, U_2, \dots, U_N\}$, 其中 N 为用户数量。域 \mathcal{X} 和 \mathcal{Y} 的物品集分别为 $I^{\mathcal{X}} = \{I_1^{\mathcal{X}}, I_2^{\mathcal{X}}, \dots, I_M^{\mathcal{X}}\}$ 和 $I^{\mathcal{Y}} = \{I_1^{\mathcal{Y}}, I_2^{\mathcal{Y}}, \dots, I_T^{\mathcal{Y}}\}$, 其中 M 和 T 分别表示域 \mathcal{X} 和 \mathcal{Y} 中的物品数。域 \mathcal{X} 的用户-物品交互可以用矩阵 $R^{\mathcal{X}} \in \mathbb{R}^{N \times M}$ 表示, 其中的值可以是显式反馈, 如评分; 或者隐式反馈, 如点击。类似地, 域 \mathcal{Y} 的用户-物品交互用 $R^{\mathcal{Y}} \in \mathbb{R}^{N \times T}$ 表示。通常, \mathcal{X} 和 \mathcal{Y} 是非常稀疏的, 因为用户只与每个域中的一小部分物品进行交互。**CDR**的目标是提高 \mathcal{X} 域和 \mathcal{Y} 域用户的推荐准确率。与文献^[6]不同, 没有区分源域和目标域, 因为这里每个域的推荐任务都是用统一的方法执行的。

3.2 整体概述

建模跨域用户行为的联合分布是必要的，因为行为在CDR中表现出相关性。通过对联合分布进行建模，了解到更具代表性的用户偏好，这有助于预测缺失的交互。提出的ETL基于对上述联合分布的建模。因此，从联合分布开始介绍ETL。通过最大化观测的联合对数似然来建模联合分布。

设 x_i 和 y_i 分别是用户 U_i 在域 \mathcal{X} 和 \mathcal{Y} 中的行为。换句话说， x_i 和 y_i 是 $R^{\mathcal{X}}$ 和 $R^{\mathcal{Y}}$ 的行向量，设 (x_i, y_i) 为用户 U_i 的一对样本。观测值的联合对数似然由单个数据点 $\sum_{i=1}^N \log p_{\theta}(x_i, y_i)$ 的似然和组成，其中 $p_{\theta}(x_i, y_i)$ 表示概率密度函数：

$$\log p_{\theta}((x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)) = \sum_{i=1}^N \log p_{\theta}(x_i, y_i) \quad (1)$$

借用VAE^[10]中边际对数似然最大化的思想，若 z_x 和 z_y 分别是 x_i 和 y_i 的潜在因素，则 $\log p_{\theta}(x_i, y_i)$ 可表示为

$$\log p_{\theta}(x_i, y_i) = D_{KL}[q_{\phi}(z_x, z_y | x_i, y_i) || p(z_x, z_y | x_i, y_i)] + \mathcal{L}(\theta, \phi; x_i, y_i), \quad (2)$$

其中第一项是近似后验 $q_{\phi}(z_x, z_y | x_i, y_i)$ 与真后验 $p(z_x, z_y | x_i, y_i)$ 的KL散度。由于这个KL项是非负的，第二项是对数似然 $\log p_{\theta}(x_i, y_i)$ 的证据下界。根据VAE^[10]中的推导， $\mathcal{L}(\theta, \phi; x_i, y_i)$ 可以写成

$$\mathcal{L}(\theta, \phi; x_i, y_i) = \mathbb{E}_{q_{\phi}(z_x, z_y | x_i, y_i)} [\log p_{\theta}(x_i, y_i | z_x, z_y)] - D_{KL}[q_{\phi}(z_x, z_y | x_i, y_i) || p(z_x, z_y)], \quad (3)$$

式中 $p_{\theta}(x_i, y_i | z_x, z_y)$ 表示用 θ 参数化的条件分布。式(3)中的第一项表示用 x_i, y_i 编码的 z_x, z_y 重构 x_i, y_i 的联合重构损失。式(3)中的第二项表示先验正则化损失，其中 $q_{\phi}(z_x, z_y | x_i, y_i)$ 预期与先验分布 $p(z_x, z_y)$ 匹配。ETL通过双自编码器结构和对抗性学习方案实现了联合重构损失和先验正则化损失。在下面的部分中，将详细介绍这两种损失，然后是目标函数。

3.3 联合重构损失

公式(3)中的第一项 $\mathbb{E}_{q_{\phi}(z_x, z_y | x_i, y_i)} [\log p_{\theta}(x_i, y_i | z_x, z_y)]$ 由一个由 ϕ 参数化的近似后验 $q_{\phi}(z_x, z_y | x_i, y_i)$ 和一个由 θ 参数化的条件分布 $p_{\theta}(x_i, y_i | z_x, z_y)$ 组成。为了简化后验 $q_{\phi}(z_x, z_y | x_i, y_i)$ 的求解，采用平均场理论来定义近似后验，具体方法如下：

$$q_{\phi}(z_x, z_y | x_i, y_i) \stackrel{\text{def}}{=} q_{\phi_x}(z_x | x_i) q_{\phi_y}(z_y | y_i), \quad (4)$$

表示潜在编码 z_x 和 z_y 分别由 x_i 和 y_i 编码。

根据数学中的定义^[5]， z_x 与 z_y 之间的等价变换定义为 $z_x = Q^{-1}z_yP$ ，其中 Q, P 为两个可逆矩阵。注意，为了简单起见，把 Q 设为单位矩阵 I 。因此，如果将 $z_{y \rightarrow x}$ 表示为等价变换后的 z_y ，将 $z_{x \rightarrow y}$ 表示为等价变换后的 z_x ，则有 $z_{y \rightarrow x} = z_y W_x$ ， $z_{x \rightarrow y} = z_x W_y$ 和 $W_x = P, W_x W_y = I$ 。等价变换假设定义了两个相关域内用户偏好之间的数学公式。它以一种比共享用户表示更适合的方式连接跨域的偏好表示(3.6节)，从而有助于模型学习每个单一域的特定于域的特征。进一步，有以下命题：

命题1. ①:给定潜在变量 z_x, z_y , 观测值 x_i 和 y_i 是条件独立的。②:给定潜变量 z_x , 观测值 x_i 与潜变量 z_y 是条件无关的。③:给定潜变量 z_y , 观测值 y_i 与潜变量 z_x 是条件无关的。

通过这个命题和ET假设,重写了条件分布

$$\begin{aligned}
p_\theta(x_i, y_i | z_x, z_y) &\stackrel{\textcircled{1}}{=} p_{\theta_x}(x_i | z_x, z_y) p_{\theta_y}(y_i | z_x, z_y) \\
&\stackrel{\textcircled{2} \text{ and } \textcircled{3}}{=} p_{\theta_x}(x_i | z_x) p_{\theta_y}(y_i | z_y) \\
&= \sqrt{p_{\theta_x}(x_i | z_x) p_{\theta_x}(x_i | z_x)} \sqrt{p_{\theta_y}(y_i | z_y) p_{\theta_y}(y_i | z_y)} \\
&\stackrel{\text{ET assumption}}{=} \sqrt{p_{\theta_x}(x_i | z_x) p_{\theta_x}(x_i | z_y W_x)} \sqrt{p_{\theta_y}(y_i | z_y) p_{\theta_y}(y_i | z_x W_y)},
\end{aligned} \tag{5}$$

其中, θ_x 是解码器 D_x 的参数, θ_y 是解码器 θ_y 的参数。式(5)表示双自编码器结构, 其中 z_x 和 z_y 分别用于重构 x_i 和 y_i 。同时, 通过等价变换, z_x 和 z_y 可以分别生成 y_i 和 x_i , $p_{\theta_x}(x_i | z_y W_x)$ 和 $p_{\theta_y}(y_i | z_x W_y)$ 表示可以学习等效变换的跨域生成。最后, 通过这种双自编码器结构, 等效转换使 z_x 和 z_y 具有维护特定域特征和学习重叠特征的能力。

等效转换的规范:不同的转换对知识转移的影响可能不同。在ETL中, 认为CDR中的等价变换应该避免用户之间的虚假关联。因此, 这里使用了正交变换, 因为它保留了向量的内积, 也就是说, 它保留了跨域的用户相似性。根据正交变换的定义, W_x 和 W_y 满足 $W_x = W$ 和 $W_y = W^T$, 其中 $W \in \mathbb{R}^{d \times d}$ 为可训练的正交映射矩阵。

综上所述, 将式(3)中的 $\mathbb{E}_{q_\phi(z_x, z_y | x_i, y_i)} [\log p_\theta(x_i, y_i | z_x, z_y)]$ 改写为

$$\begin{aligned}
\mathbb{B}_{q_\phi(z_x, z_y | x_i, y_i)} [\log p_\theta(x_i, y_i | z_x, z_y)] &= \frac{1}{2} \left\{ \mathbb{E}_{q_{\phi_x}(z_x | x_i)} [\log p_{\theta_x}(x_i | z_x)] + \mathbb{E}_{q_{\phi_y}(z_y | y_i)} [\log p_{\theta_x}(x_i | z_y W)] \right. \\
&\quad \left. + \mathbb{E}_{q_{\phi_y}(z_y | y_i)} [\log p_{\theta_y}(y_i | z_y)] + \mathbb{B}_{q_{\phi_x}(z_x | x_i)} [\log p_{\theta_y}(y_i | z_x W^T)] \right\}.
\end{aligned} \tag{6}$$

式(6)的最优值与式(6)的最优值乘以常数相同。因此, 为简化表达式, 可将重构损失记为

$$\begin{aligned}
\min_{\phi_x, \phi_y, \theta_x, \theta_y, W} \mathcal{L}_{JRL} &= -\mathbb{E}_{q_{\phi_x}(z_x | x_i)} [\log p_{\theta_x}(x_i | z_x)] \\
&\quad - \mathbb{E}_{q_{\phi_y}(z_y | y_i)} [\log p_{\theta_x}(x_i | z_y W)] \\
&\quad - \mathbb{E}_{q_{\phi_y}(z_y | y_i)} [\log p_{\theta_y}(y_i | z_y)] \\
&\quad - \mathbb{E}_{q_{\phi_x}(z_x | x_i)} [\log p_{\theta_y}(y_i | z_x W^T)] \\
&\quad + \lambda \left(\|z_x - z_x W^T W\|_F^1 + \|z_y - z_y W W^T\|_F^1 \right),
\end{aligned} \tag{7}$$

其中最后一项表示当变换指定为正交映射时等价变换的正则化损失。 λ 是衡量正则化重要性的超参数。 $\{\phi_x, \phi_y\}$ 和 $\{\theta_x, \theta_y\}$ 分别是编码器和解码器的参数。

3.4 先验正则化损失

第二项 $D_{KL}[q_\phi(z_x, z_y | x_i, y_i) \| p(z_x, z_y)]$ 涉及联合先验 $p(z_x, z_y)$, 它表示 z_x, z_y 的复杂先验。在本文中, 为了简单起见, 设置 $p(z_x, z_y) = p(z_x) p(z_y)$ 考虑式(4), 先验正则化项为

$$\begin{aligned}
D_{KL}[q_\phi(z_x, z_y | x_i, y_i) \| p(z_x, z_y)] &= D_{KL}[q_{\phi_x}(z_x | x_i) \| p(z_x)] \\
&\quad + D_{KL}[q_{\phi_y}(z_y | y_i) \| p(z_y)],
\end{aligned} \tag{8}$$

其中 $p(z_x)$ 和 $p(z_y)$ 是先验分布。公式(8)描述了将 $q_{\phi_x}(z_x | x_i)$ 匹配到先验 $p(z_x)$ 并将 $q_{\phi_y}(z_y | y_i)$ 匹配到先验 $p(z_y)$ 的正则化。

由于在KL-散度中不容易推导出一些复杂先验的显式公式，因此ETL采用了对抗性分布匹配，可以在没有硬推导的情况下为潜在代码施加任意先验分布。也就是说，利用对抗性学习来执行分布匹配 $q_{\phi_x}(z_x | x_i)$ 和 $p(z_x)$ 以及 $q_{\phi_y}(z_y | y_i)$ 和 $p(z_y)$ 。先验正则化损失可以用以下对抗性学习方案：

$$\begin{aligned} \min_{\psi} \max_{\phi} \mathcal{L}_{PRL} = & -\mathbb{E}_{z_x \sim p(z_x)} [\log \mathcal{D}_x(z_x)] \\ & -\mathbb{E}_{z_x \sim q_{\phi_x}(z_x | x_i)} [\log (1 - \mathcal{D}_x(z_x))] \\ & -\mathbb{E}_{z_y \sim p(z_y)} [\log \mathcal{D}_y(z_y)] \\ & -\mathbb{E}_{z_y \sim q_{\phi_y}(z_y | y_i)} [\log (1 - \mathcal{D}_y(z_y))], \end{aligned} \quad (9)$$

其中 $\phi = \{\phi_x, \phi_y\}$ 与式(7)中的定义相同， $\psi = \{\psi_x, \psi_y\}$ 是鉴别器 $\mathcal{D}_x, \mathcal{D}_y$ 的参数。 $p(z_x)$ 和 $p(z_y)$ 分别是 z_x 和 z_y 的先验分布。

与式(8)中的KL-散度相比，式(9)中采用的对抗性分布匹配有几个优点。KL-散度试图将 $q_{\phi_x}(z_x | x_i)$ 匹配到先前的 $p(z)$ ，这将有丢失输入 x_i 信息的风险。相反，潜在空间中的对抗性分布匹配使得后验 $q_{\phi_x}(z_x | x_i)$ 成为聚合后验 $q_{\phi_x}(z_x)$ ，这促使 z_x 匹配 $p(z_x)$ 的整个分布。因此， z_y 可以以类似的方式匹配 $p(z_y)$ 的整个分布。同时，可以避免对抗学习中的模式崩溃问题，因为涉及重建损失，这有助于潜在嵌入匹配先验和整个真实数据分布。

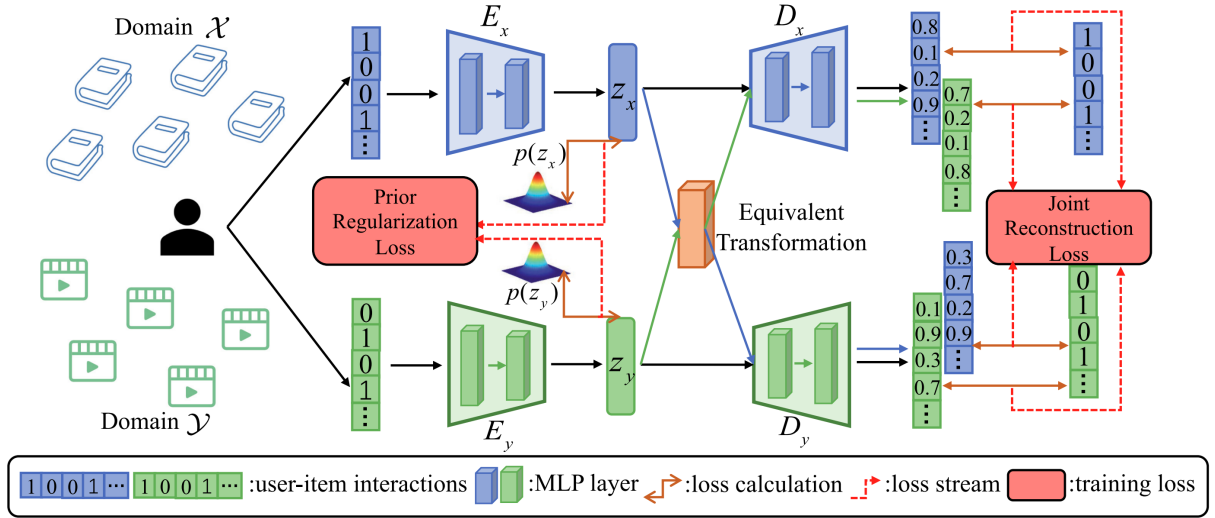


图 1. ETL示意图（来源于原论文）

3.5 目标函数

结合式(7)和式(9)，可以通过优化以下目标函数实现式(3)中ELBO的最大化：

$$\min_{\Theta} \max_{\Phi} \mathcal{L}_{ETL} = \mathcal{L}_{JRL} + \eta \mathcal{L}_{PRL}, \quad (10)$$

其中 $\Theta = \{\phi_x, \phi_y, \theta_x, \theta_y, \psi_x, \psi_y, W\}$ 和 $\Phi = \{\phi_x, \phi_y\}$ 为网络参数。 η 是衡量先验正则化项重要性的超参数。

ETL的架构如图1所示，其中 E_x 是 $q_{\phi_x}(z_x | x_i)$ 的编码器， E_y 是 $q_{\phi_y}(z_y | y_i)$ 的编码器。同样， D_x 和 D_y 分别是 $p_{\theta_x}(x_i | \cdot)$ 和 $p_{\theta_y}(y_i | \cdot)$ 的解码器。鉴别器 \mathcal{D}_x 和 \mathcal{D}_y 是为对抗学习设计的。在实现中，所有的 $E_x, E_y, D_x, D_y, \mathcal{D}_x, \mathcal{D}_y$ 都是两层MLP，其中Relu作为非线性激活函数。使用标准高斯分布，如 $p(z_x) \sim \mathcal{N}(0, 1)$ 和 $p(z_y) \sim \mathcal{N}(0, 1)$ ，这是基于对抗性学习的方法中的常见选择。虽然 $p(z_x)$ 和 $p(z_y)$ 都遵循标准高斯分布，但这并不意味着 z_x 和 z_y 在同一个潜在空间中，不会破坏在CDR中对用户偏好的两个特征建模的动机。此外，当用户-物品交互是显式反馈时，预测与真实数据之间的重建损失是MSE，当用户-物品交互是隐式反馈时，预测与真实数据之间的重建损失是二元交叉熵。

3.6 等效变换的讨论

在本节中，对等效转换进行一些讨论，以深入了解ETL如何连接到以前的方法并与以前的方法有所不同。尽管采用共享用户表示的方法没有提到这一点，但他们试图对用户行为跨域的联合分布进行建模。根据耦合合理，如果要对两个边际观测值 \mathcal{X}, \mathcal{Y} 的联合分布进行建模，则必须有一个假设来描述对应的潜在因子 z_x 和 z_y 之间的关系。假设可以表示为 $z_x = z_y W_x, z_y = z_x W_y$ (或非线性形式)。有两种可能的表述，无约束和有约束。无约束的公式化意味着 W_x 和 W_y 之间不存在约束，类似于使用MLP作为跨域映射函数的思想。而受约束的则对 W_x 和 W_y 施加约束。

在CDR中，不同域的 z_x 和 z_y 具有相关性。因此，在ETL中，采用约束公式，具体来说就是ET假设。这个假设涉及到 $W_x W_y = I$ 的等价变换，当 $W_x = W_y = I$ 时，共享用户表示的思想。然而， $W_x = W_y = I$ 会使特定域的特征被抑制。相反，具有ET假设的ETL允许用户在不同域的偏好具有学习特定于域的特征以及重叠特征的能力。ET假设确保了跨域的生成，从而促进了更好的知识转移。无约束和有约束公式的不同变体将通过实验进行讨论。可能存在两个用户行为没有关系或几乎没有关系的域，因此不适合进行知识转移和应用等效的基于转换的模型。判断ETL模型是否可以在两个域中工作的一种直观的方法是看两个域是否不同，同时又具有可以连接用户行为的共同属性。例如，Movie和Book域是两个不同的域，并且具有明显的重叠属性，例如主题(爱情、战争等)。喜欢爱情电影的用户阅读言情小说的概率很高。

4 实验与分析

在本节中，首先给出有关数据集和实验设置的详细信息。然后，通过在多个公共数据集上与最近最先进的方法进行比较，对ETL进行评估。

4.1 数据集描述

为了评估ETL的有效性，使用来自Amazon的三个数据集对ETL进行评估，这三个数据集是Movies and TV(Movie), Books(Book), 以及CDs and Vinyl(Music)。在三个数据集之间进行两两组合，找到两个域中的共享用户。也就是说，获得Movie & Book, Movie & Music, Music & Book作为实验数据集。与显式反馈(例如，用户对物品的评分)相比，隐式反馈(例如，用户点击或不点击某个物品)在现实世界的推荐系统中更为常见，因此，将重点放在隐式用户-物品

交互上。换句话说，用户-物品交互矩阵是二进制矩阵，如果用户与物品交互，则值为1(观察到或单击)，否则为0(未观察到或未单击)。由于这些数据集中的用户-物品交互是0到5之间的评分，因此将3,4,5的评分转换为正样本。最后，过滤交互次数少于5的用户和物品。数据集统计信息如表2所示。如表2所示，每个数据集中的两个域都非常稀疏，至少99.86的交互未被观察到。这对大多数基于聚类和基于MF的CDR方法提出了巨大的挑战，因为这些方法通常需要至少一个域中的密集交互^[16]。此外，Movie & Book物品数量不平衡，Movie & Music物品密度不平衡，这为不同的CDR方法提供了更全面的评价条件。

表 2. 数据集统计

数据集	Movie & Book		Movie & Music		Music & Book	
用户数	29476		15914		16267	
域名	Movie	Book	Movie	Music	Music	Book
物品数	24091	41884	17794	20058	18467	23988
交互数	591258	579131	416228	280398	233251	291325
密度	0.08%	0.05%	0.14%	0.09%	0.08%	0.07%

4.2 实验设置

性能评估。在物品推荐中，leave-one-out(LOO)评价法被广泛使用，本文也通过LOO进行评价。换句话说，为每个用户随机保留两个物品，一个作为验证物品，另一个作为测试物品。随机抽取99个没有被用户交互的物品作为负面物品，然后评估用户如何根据负面物品对验证和测试物品进行排序。由于关注推荐中的隐式反馈，采用了三种广泛使用的评估指标:命中率(HR)、归一化折损累积增益(NDCG)和平均倒数排名(MRR)，值越高意味着这三个指标的推荐性能越好。预测的排名 $\text{topK} = 5, 10$ 。此外，在训练过程中，根据验证集上的性能保存最佳训练模型，并使用保存的模型进行测试。

参数设置。复现的设备的CUDA Version:11.2，使用Pytorch 1.10.2和Python 3.7.12。所有方法的embedding维度大小都固定为200。使用Adam优化器对所有模型进行优化^[9]，并将批大小设置为256。默认的Xavier初始化器^[3]用于初始化所有模型参数。对于所有的方法，学习率为0.001。训练epoch的个数设置为300，可以保证所有模型的收敛性。在ETL中，为了简单起见，没有调优超参数 η ，而是在三个数据集测试中将其固定为1.0。对于超参数 η ，根据验证集上的性能在[0.1,0.5,1.0,2.0,5.0,10.0]之间进行调优。然后，得到Movie & Book的 $\lambda = 5.0$ ，Movie & Music的 $\lambda = 0.5$ ，Music & Book的 $\lambda = 1.0$ 。

5 实验结果分析

实验结果如下：

表 3. topK=5 结果

	Movie & Book					
topK	topK=5					
Domain	Movie			Book		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.5114	0.3829	0.3410	0.5093	0.4026	0.3672
	Movie & Music					
topK	topK=5					
Domain	Movie			Music		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.4819	0.3590	0.3176	0.5227	0.4000	0.3597
	Music & Book					
topK	topK=5					
Domain	Music			Book		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.4629	0.3571	0.3217	0.4642	0.3568	0.3213

表 4. topK=10 结果

	Movie & Book					
topK	topK=10					
Domain	Movie			Book		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.6433	0.4252	0.3584	0.6301	0.4410	0.3830
	Movie & Music					
topK	topK=10					
Domain	Movie			Music		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.6220	0.4032	0.3358	0.6523	0.4418	0.3770
	Music & Book					
topK	topK=10					
Domain	Music			Book		
Metrics	HR	NDCG	MRR	HR	NDCG	MRR
ETL	0.5853	0.3966	0.3378	0.5771	0.3932	0.3363

三个数据集不同潜在向量维度，实验结果如下：

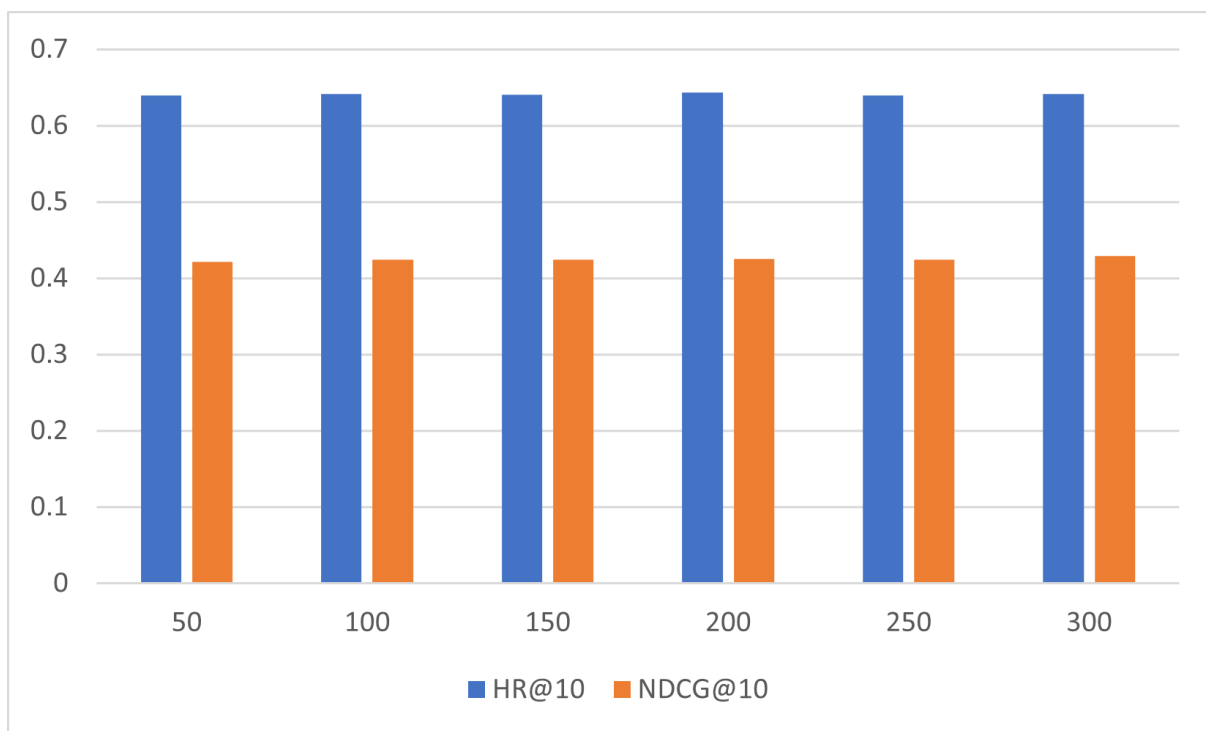


图 2. Movie & Book-Movie

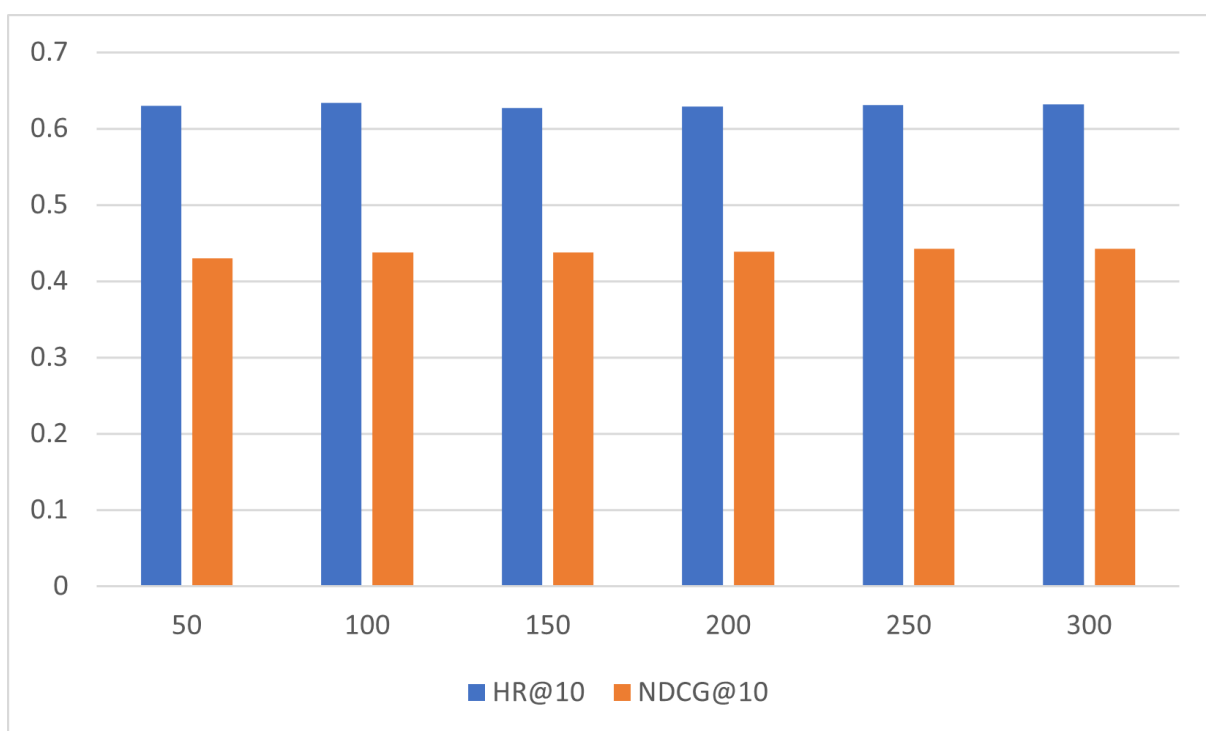


图 3. Movie & Book-Book

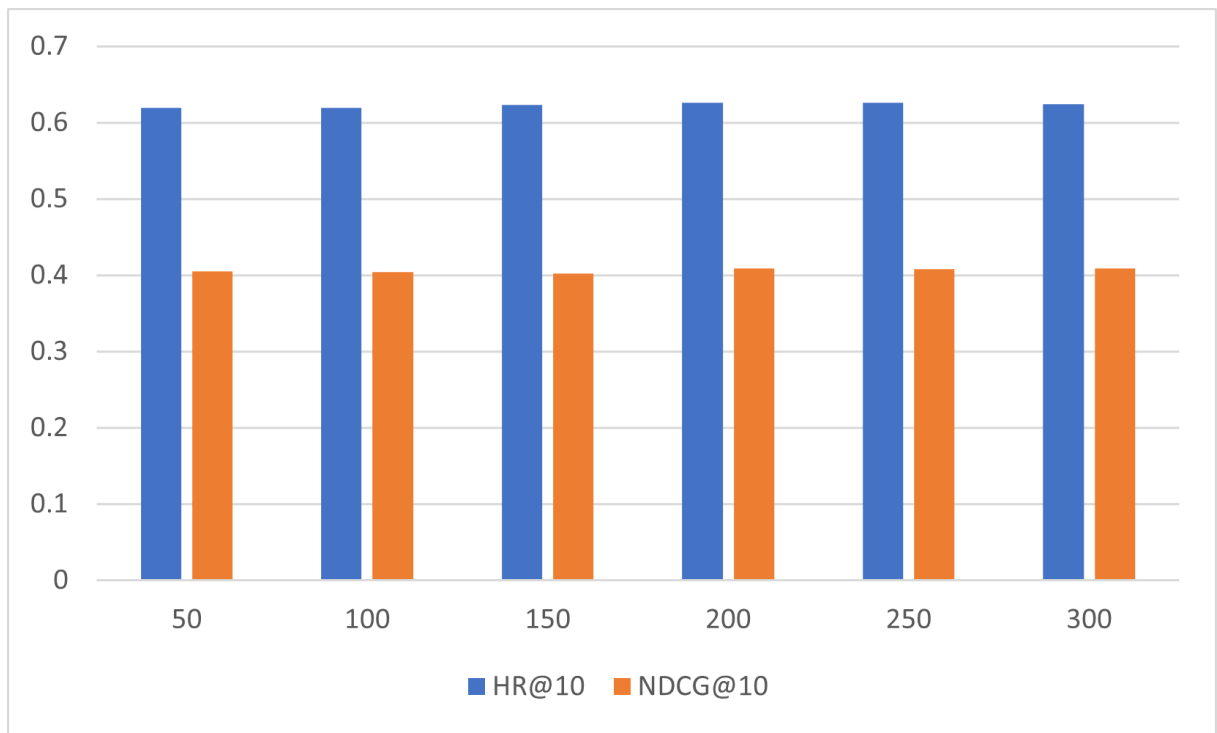


图 4. Movie & Music-Movie

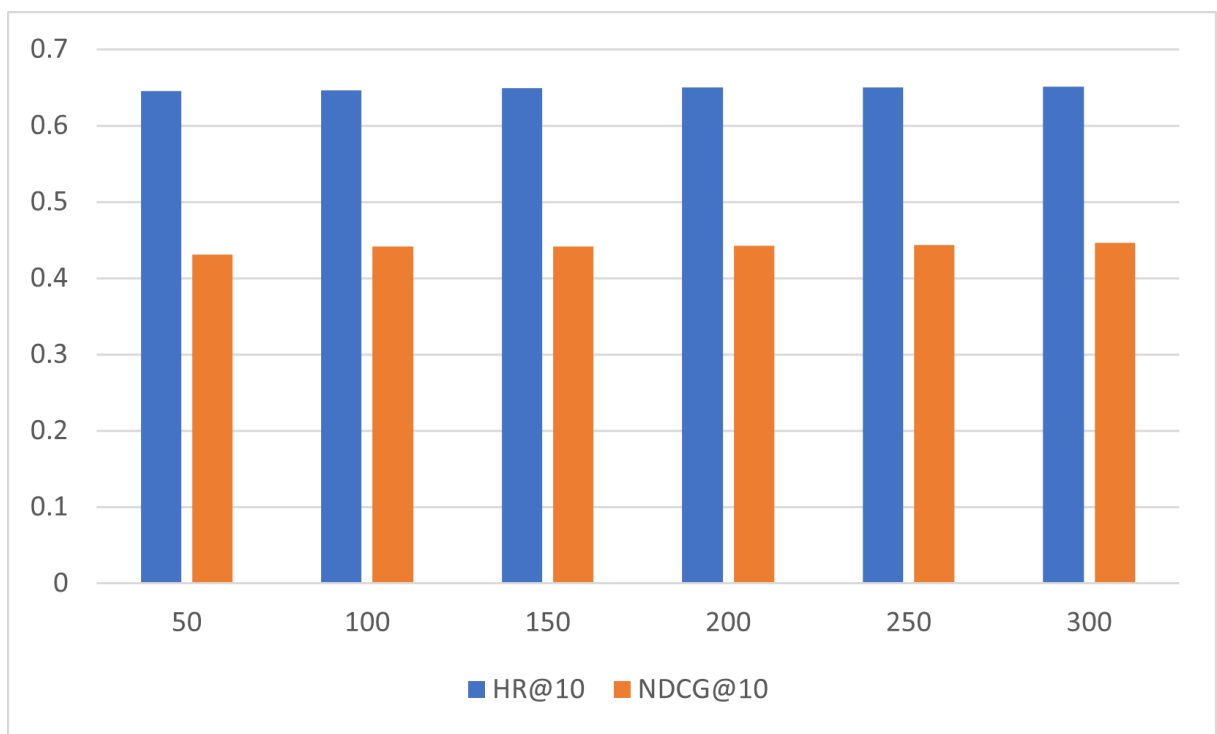


图 5. Movie & Music-Music

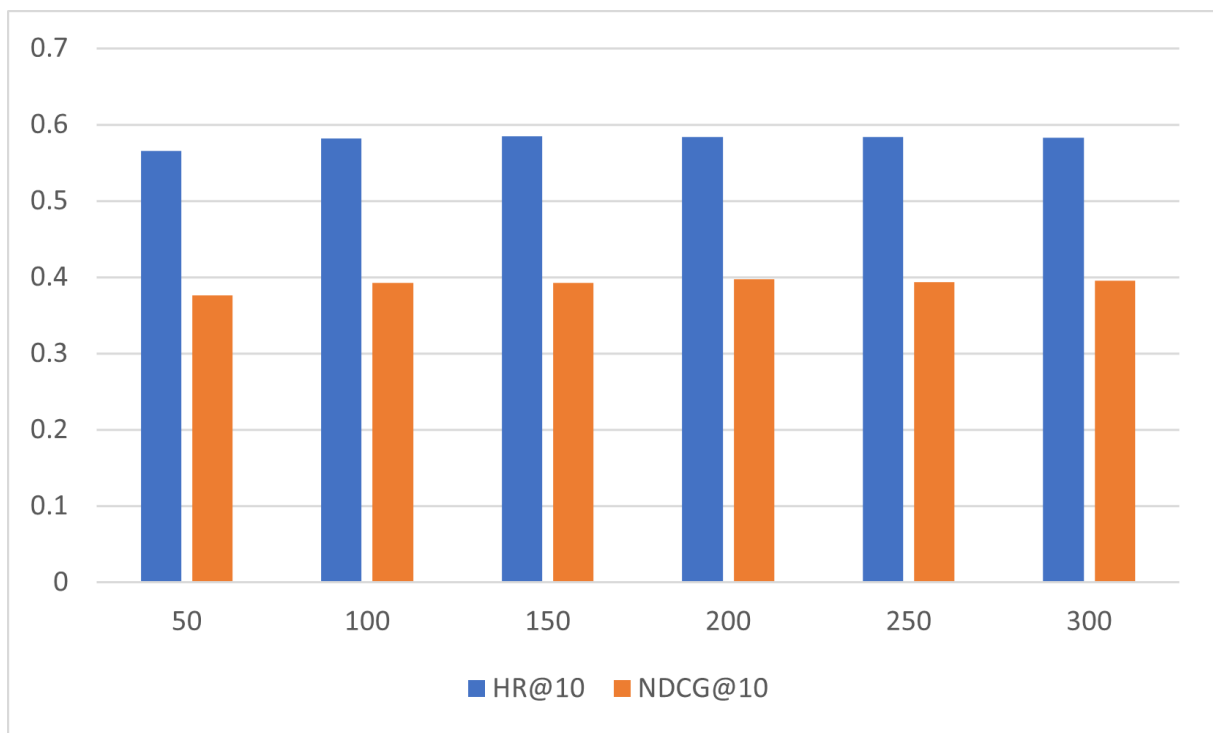


图 6. Music & Book-Music

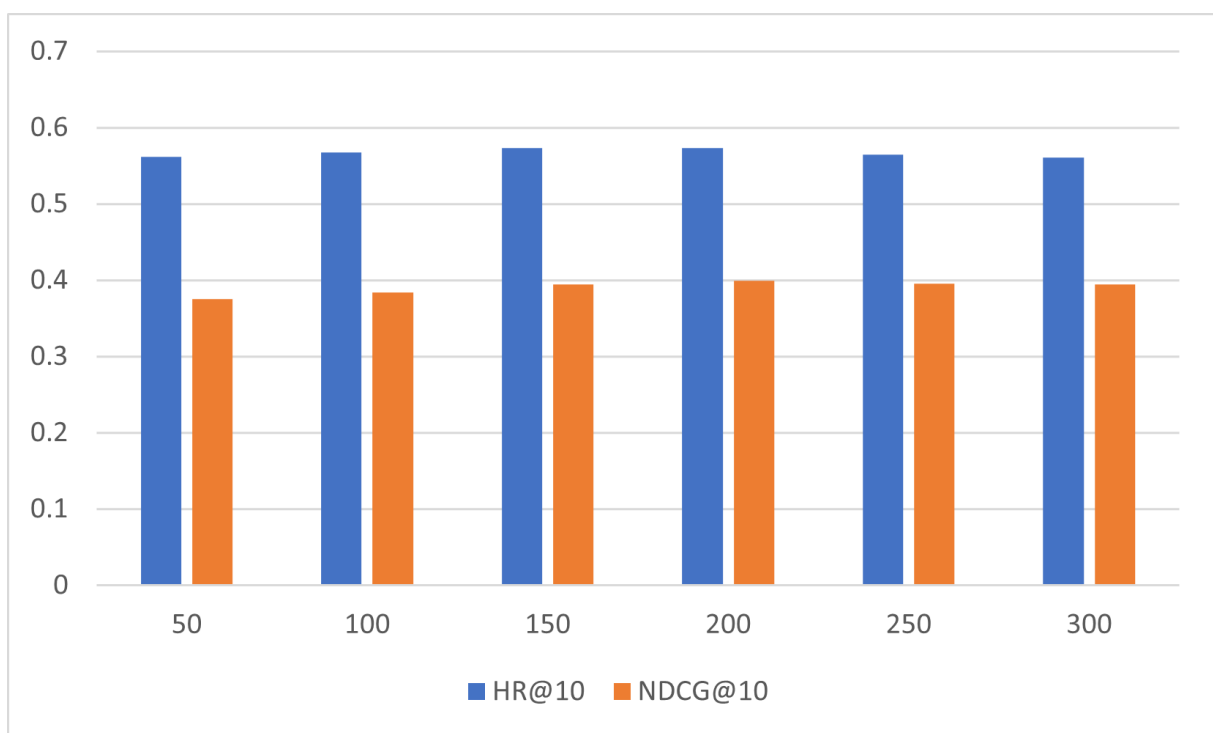


图 7. Music & Book-Book

不同转换形式的结果如图所示，其中Trans0是使用的正交变换，Trans1，2，3 在表5中定义的其他三种不同的变换。

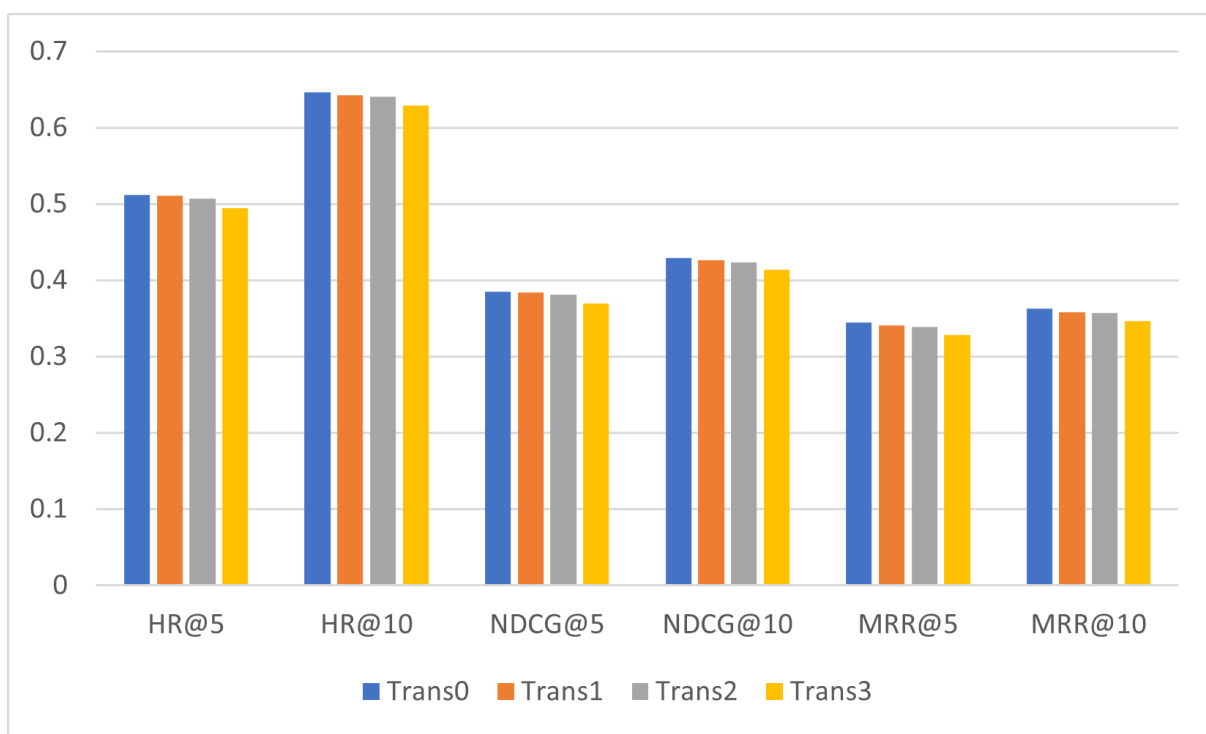


图 8. Movie & Book-Movie不同的转换方式对比

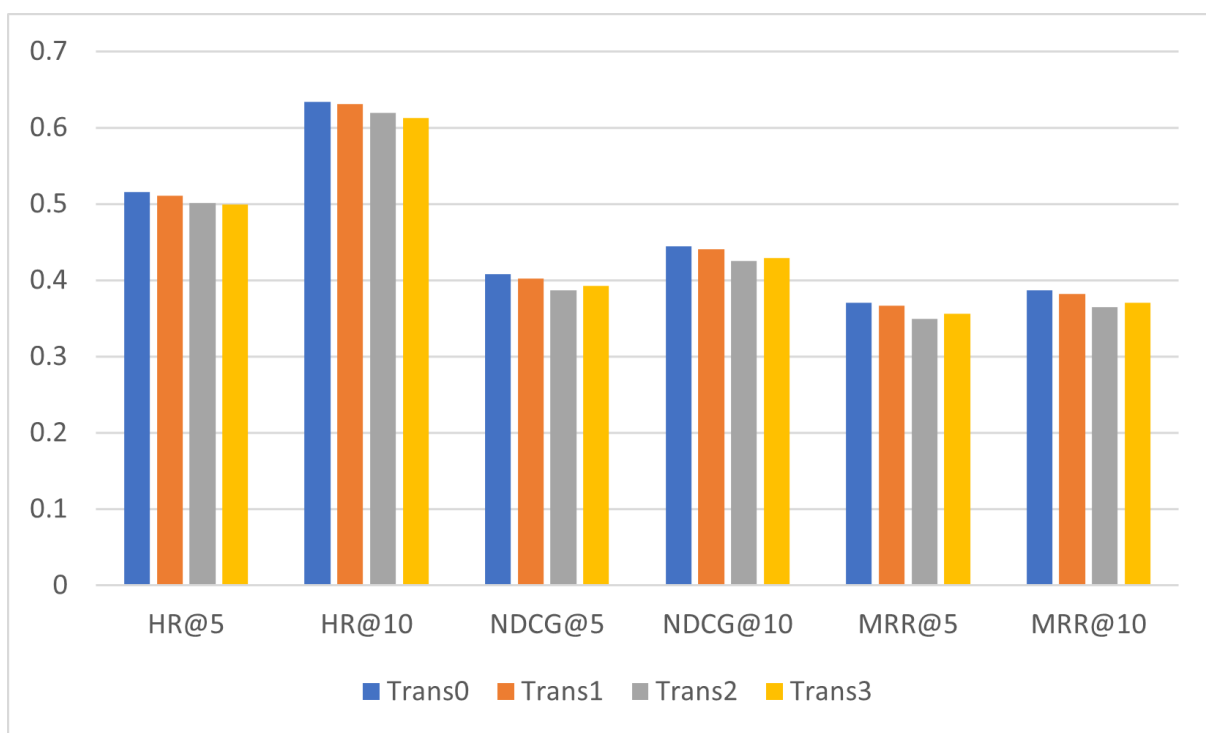


图 9. Movie & Book-Book不同的转换方式对比

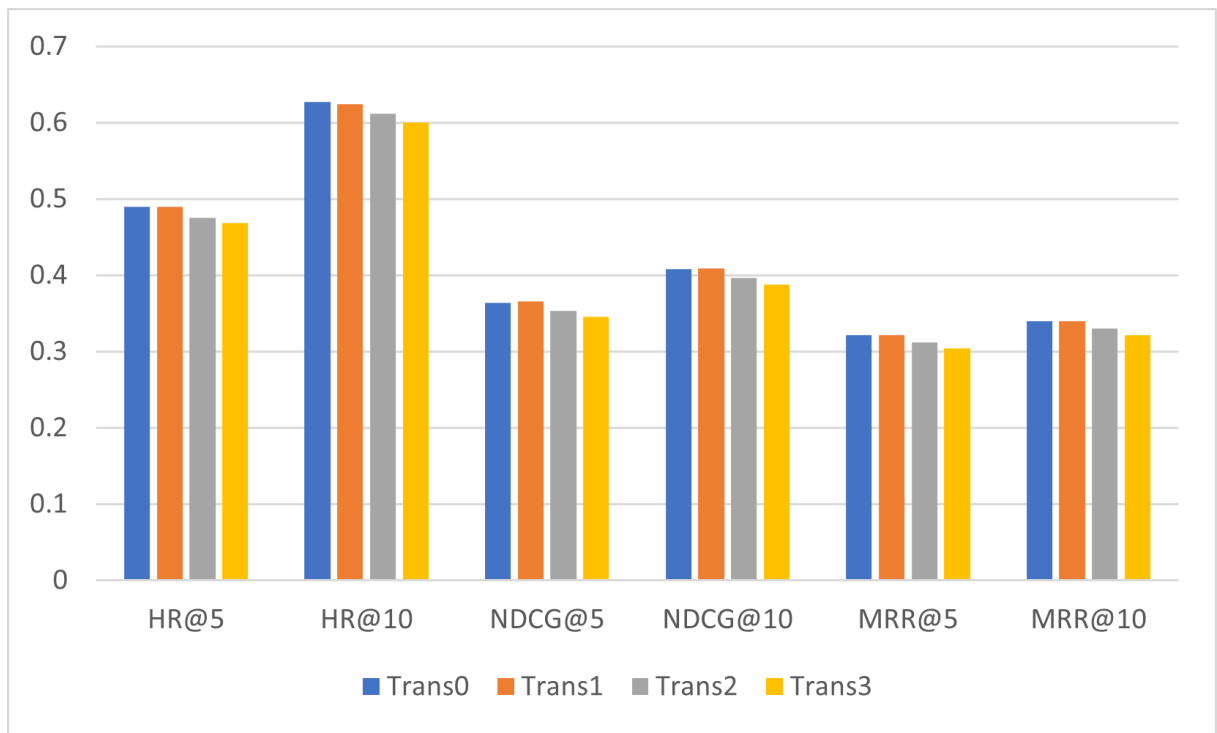


图 10. Movie & Music-Movie不同的转换方式对比

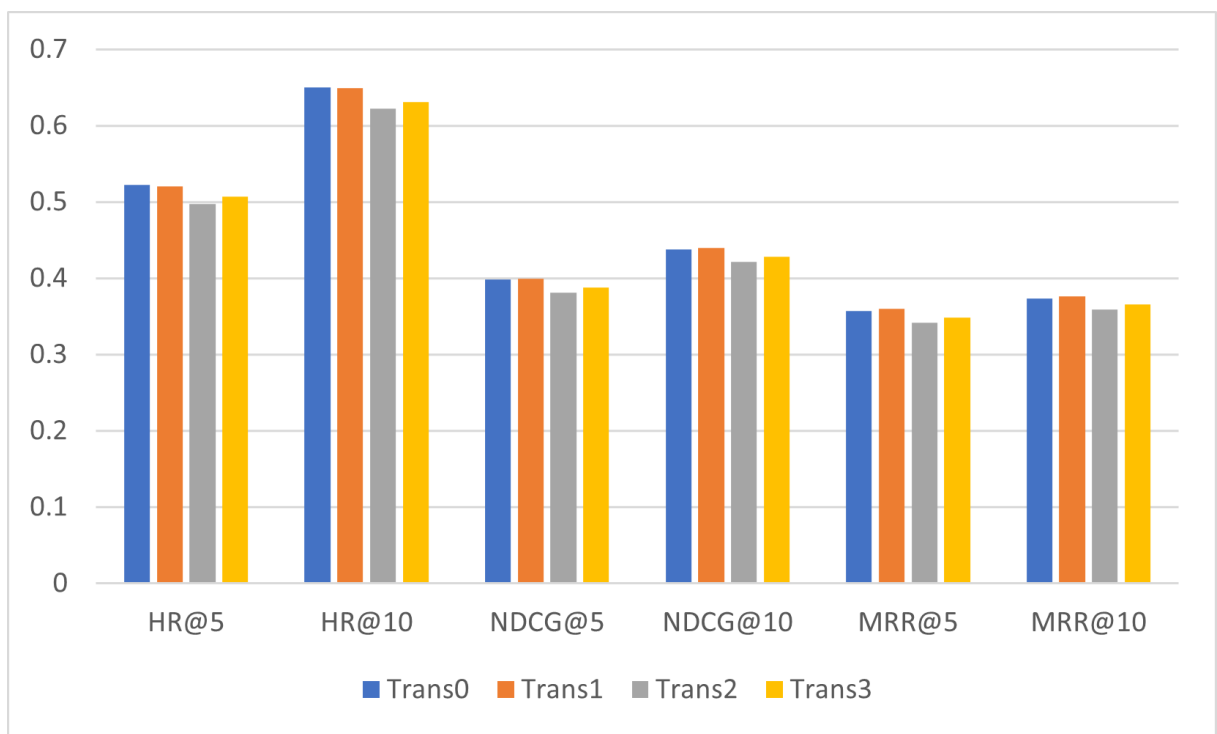


图 11. Movie & Music-Music不同的转换方式对比

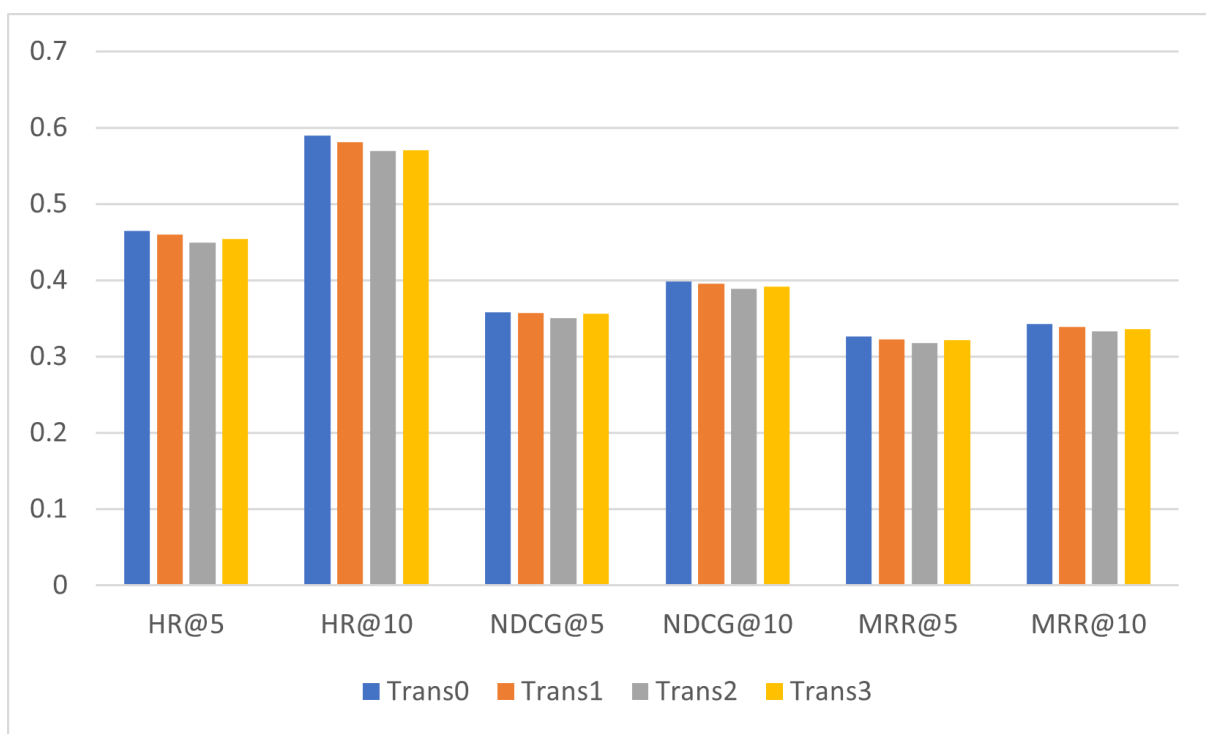


图 12. Music & Book-Music不同的转换方式对比

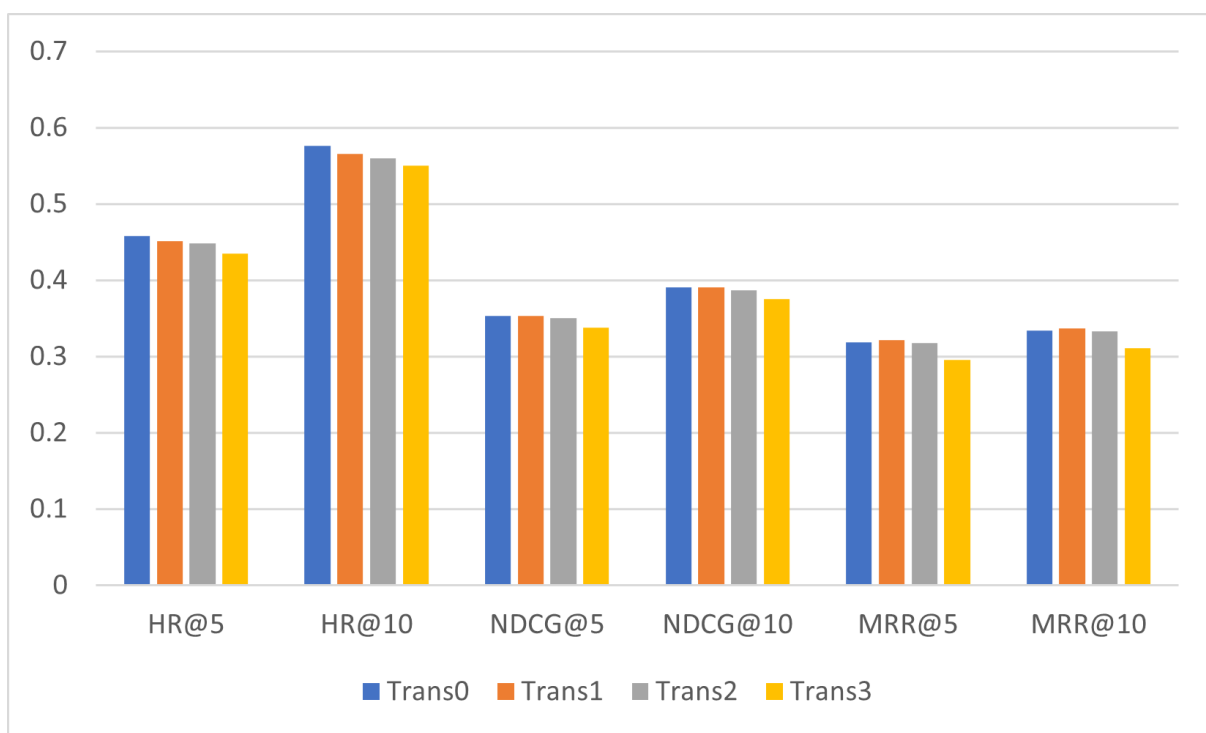


图 13. Music & Book-Book不同的转换方式对比

表 5. 不同转换方式定义

Trans	Formulation	Type
Trans1	$z_{y \rightarrow x} = z_y W_x, z_{x \rightarrow y} = z_x W_y$	unconstrained (linear)
Trans2	$z_{y \rightarrow x} = \sigma(z_y W_x^1) W_x^2, z_{x \rightarrow y} = \sigma(z_x W_y^1) W_y^2$	unconstrained (non-linear)
Trans3	$z_{y \rightarrow x} = z_y W_x, z_{x \rightarrow y} = z_x W_y$ s.t. $\min \ z_x - z_{x \rightarrow y} W_x\ _F^1$ s.t. $\min \ z_y - z_{y \rightarrow x} W_y\ _F^1$	constrained (linear, equivalent)

Trans1和Trans2都不是等价变换。Trans3是一个简单的等价变换。值得注意的是，等效转换比非等效转换(Trans1,2)表现出更好的性能。这验证了我们的ET假设的正确性，该假设有利于ETL更好地学习用户偏好的覆盖。而不受约束的推荐则会降低推荐的性能，因为它们过于灵活并且在推荐中很容易对稀疏数据进行过拟合。本复现报告在三个公共数据集上进行的原始实验、不同潜在向量维度的实验以及不同的转换方式实验，结果与原论文报告结果一致。

正交变换(Trans0)优于其他3种变换。这是合理的，因为只有Trans0是等效的，并且不会在知识转移后引入用户之间的虚假相关性。

6 总结与展望

在复现的这个论文中，他们提出了一种基于联合分布匹配的方法，称为ETL，该方法同时捕获CDR的重叠特征和特定域的特征。ETL需要跨域等价转换假设，更好地进行用户偏好建模和用户行为预测。只学习一种特征对CDR模型的性能是有限的。一个执行良好的CDR模型应该在捕获重叠特征的基础上捕获特定域的特征。提出的ETL通过最大化域联合用户行为的联合可能性，并且不需要启发式的方法来选择训练用户或设计网络架构。本复现报告在三个公共数据集上进行的原始实验、不同潜在向量维度的实验以及不同的转换方式实验，结果与原论文报告结果一致。

虽然ETL表现出更好的性能，但是仍然有一些不足限制了它的潜力。首先，在论文中，对联合先验 $p(z_x, z_y)$ 做了一个简化的高斯假设。提供更多信息性先验知识的复杂先验可以稍后探索。其次，在公共CDR中，通常会考虑推荐流行的相关域。在未来，研究如何在不受欢迎的域之间进行推荐是一个值得关注的问题。第三，如何将ETL与辅助信息相结合来提高推荐性能，甚至解决冷启动问题也是一个值得关注的问题。

参考文献

- [1] Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. Cross-domain recommender systems. *Recommender systems handbook*, pages 919–959, 2015.
- [2] Xu Chen, Ya Zhang, Ivor W Tsang, Yuangang Pan, and Jingchao Su. Toward equivalent transformation of user preferences in cross domain recommendation. *ACM Transactions on Information Systems*, 41(1):1–31, 2023.
- [3] 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.
- [4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [5] William H Gustafson and Julius M Zelmanowitz. On matrix equivalence and matrix equations. *Linear Algebra and Its Applications*, 27:219–224, 1979.
- [6] Ming He, Jiuling Zhang, Peng Yang, and Kaisheng Yao. Robust transfer learning for cross-domain collaborative filtering using multiple rating patterns approximation. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 225–233, 2018.
- [7] Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 2013.
- [8] Muhammad Murad Khan, Roliana Ibrahim, and Imran Ghani. Cross domain recommender systems: a systematic literature review. *ACM Computing Surveys (CSUR)*, 50(3):1–34, 2017.
- [9] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [10] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [11] Bin Li, Qiang Yang, and Xiangyang Xue. Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th annual international conference on machine learning*, pages 617–624, 2009.
- [12] Jin Li, Zhaohui Peng, Senzhang Wang, Xiaokang Xu, Philip S Yu, and Zhenyun Hao. Heterogeneous graph embedding for cross-domain recommendation through adversarial learning. In *Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020, Jeju, South Korea, September 24–27, 2020, Proceedings, Part III* 25, pages 507–522. Springer, 2020.

- [13] 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.
- [14] Chang-Dong Wang, Yan-Hui Chen, Wu-Dong Xi, Ling Huang, and Guangqiang Xie. Cross-domain explicit–implicit–mixed collaborative filtering neural network. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(11):6983–6997, 2021.
- [15] Tianxin Wang, Fuzhen Zhuang, Zhiqiang Zhang, Daixin Wang, Jun Zhou, and Qing He. Low-dimensional alignment for cross-domain recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3508–3512, 2021.
- [16] Feng Yuan, Lina Yao, and Boualem Benatallah. Darec: Deep domain adaptation for cross-domain recommendation via transferring rating patterns. *arXiv preprint arXiv:1905.10760*, 2019.
- [17] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 793–803, 2019.
- [18] Shengjia Zhao, Jiaming Song, and Stefano Ermon. Infovae: Information maximizing variational autoencoders. *arXiv preprint arXiv:1706.02262*, 2017.
- [19] 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.