

赞同 20

分享

蚂蚁2024：开放世界视角下的跨领域序列推荐



SmartMindAI

专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注

已关注

20 人赞同了该文章

Introduction

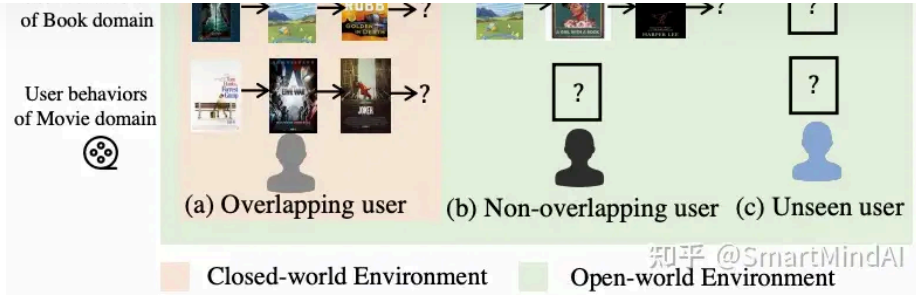
单域序列推荐 (SDSR) 在电子商务[推荐系统](#)⁺中广受关注，因为它能根据用户的交互序列动态捕捉其偏好。然而，传统的序列推荐方法往往受限于数据稀疏性和新用户冷启动问题，这些问题严重影响了它们的推荐准确性。

为克服SDSR中的这些问题，AMID（适应性多兴趣偏移矫正框架）应运而生，它由MIM（多兴趣信息模块⁺）和DRE（双重稳健估计器）构成，特别针对开放世界环境设计，能提升非重叠用户在CDSR（跨域序列推荐）中的表现。MIM通过整合各类用户兴趣，帮助深入理解用户潜在需求；DRE则用于校正跨域推荐可能产生的偏差。实验基于MYbank-CDR数据集，在[亚马逊](#)⁺和MYbank等场景验证了AMID的有效性，并在支付宝进行了实际的A/B测试以验证其在开放环境下的应用。

Motivation: Towards Open-World Cross-Domain Recommendation

我们针对开放世界环境进行研究，通过在亚马逊数据集上比较了单领域和跨领域序列推荐模型（BERT4Rec, GRU4Rec, SASRec, Pi-Net, DASL, C²DSR），着重考察了SDSR在不同重叠情况下的表现。在电影领域，SASRec作为SDSR表现出最佳效果，但在100%重叠的音乐领域，即使在CDSR（DASL）下，SASRec仍有优势。这个现象揭示了CDSR方法在部分重叠条件下性能下降的问题，即它们过于依赖重叠用户和跨域信息，这是**第一个挑战**。

知乎



令人意外的是，SDSR模型对重叠用户的需求似乎并不那么依赖，反而在低重叠情况下，某些SDSR方法在高重叠比例下表现甚至超过CDSR。这违背了传统认知，验证了需要设计针对开放世界环境的高性能CDSR模型的需求。我们的实验结果表明，AMID正是为解决这一问题而设计，展示了其在实际应用中的有效性。

在开放世界中，多领域数据选择偏见是个关键问题，因为我们面对的是非重叠用户为主的场景。AMID框架正是为解决这个问题而提出，通过设计一个多兴趣信息模块和双重稳健估计器，既能处理各类用户，又能有效地消除数据分布变化带来的选择偏见。

Problem Definition

在论文中，我们分析了一个部分重叠的多领域协同序列推荐（CDSR）问题，其中每个领域子集为 $\mathcal{Z} = \{\mathcal{Z}_1, \dots, \mathcal{Z}_{|\mathcal{Z}|}\}$

用户集合 $\mathcal{U} = \{u_1, \dots, u_{|\mathcal{U}|}\}$

项目集合 $\mathcal{V} = \{v_1^{\mathcal{Z}_1}, \dots, v_{|\mathcal{V}|}^{\mathcal{Z}_{|\mathcal{Z}|}}\}$

其中，非重叠用户仅在部分领域有行为记录，而重叠用户则有多领域历史。用户的行为序列 $\mathcal{S} = \{S^{\mathcal{Z}_1}, \dots, S^{\mathcal{Z}_{|\mathcal{Z}|}}\}$

表示各个领域的活动。数据集 \mathcal{D} 是用户行为的交叉矩阵，我们旨在构建一个能够考虑序列依赖关系的个性化推荐模型。传统的跨领域推荐（CDR）通常不考虑这种行为间的关联，而CDSR则不然，它聚焦于捕捉这些动态关系。CDR的目标可以数学上表示为：

$$\min_f \sum_{u,i} (v_{T+1}^{\mathcal{Z}_i}(u) - f(u, S^{\mathcal{Z}_{1:i-1}}))^2$$

这里 $f(u, S^{\mathcal{Z}_{1:i-1}})$ 是预测函数⁺，它根据用户在前 $i-1$ 领域的序列信息来估计用户在领域 \mathcal{Z}_i 中的下一个项目 $v_{T+1}^{\mathcal{Z}_i}(u)$ 。通过这样的方法，CDSR试图减少在开放世界环境下由于数据选择偏见造成的推荐不准确。

$$\operatorname{argmax} P^X (r_{u,v}^X = v | U^X, U^Y, V^X), \text{ if } v \in \mathcal{V}^X.$$

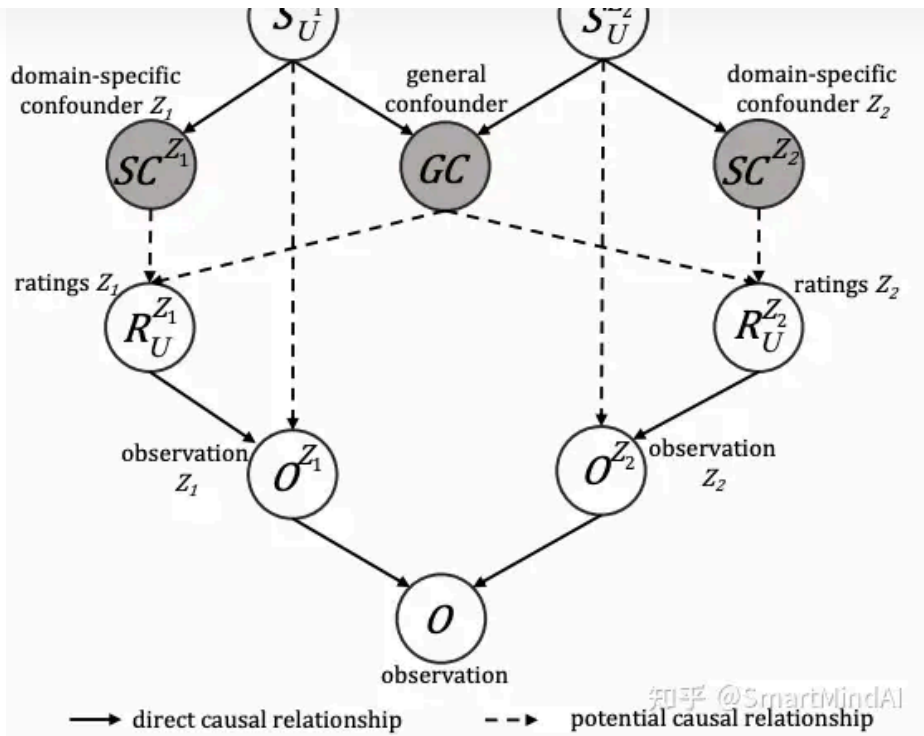
$$\text{CDSR目标} = \min_f \sum_u \sum_v (r_{u,v}^Y - f(u, S_u^Y))^2$$

其中 $f(u, S_u^Y)$ 是通过用户在领域 Y 的所有历史序列 S_u^Y 来个性化预测用户对下一个项目的兴趣。这个优化过程旨在降低因领域间数据不完整导致的推荐误差。

$$\operatorname{argmax} P^X (S_{|S^X|+1}^X = v | S^X, S^Y, U^X, U^Y, V^X), \text{ if } v \in \mathcal{V}^X.$$

Causal Graph

知乎

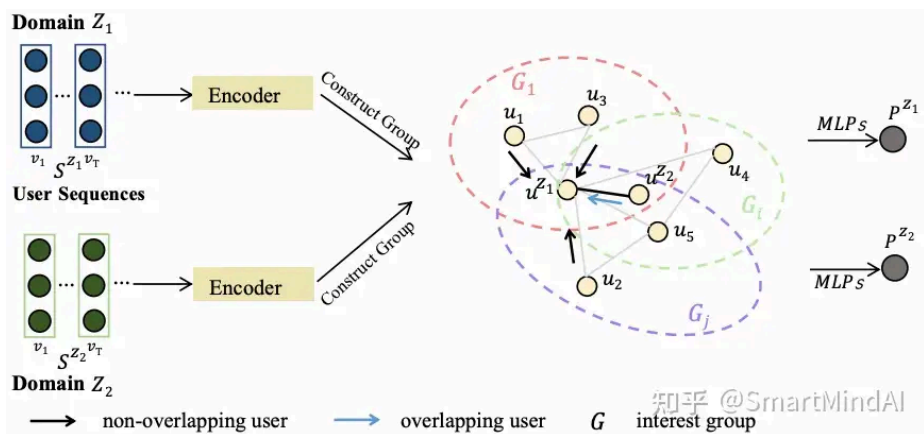


为克服开放世界环境下多领域数据不完整的问题，我们引入因果框架AMID，通过处理 S_U^Z （用户历史事件） R_U^Z （观察到评分） SC^Z （特定领域混杂因子）和 GC （一般混杂因子）。 O 是基于 O^{Z_1} 和 O^{Z_2} 的观测变量，若用户在两个领域都未被充分观察，则 $O = 0$ ，否则 $O = 1$ 。SDSR专注于学习 SC^Z ，而CDSR基于共享用户估计 GC 。两者分别关注特定领域和普遍影响。然而，不平衡的观测数据，如 S_U^Z 和 R_U^Z 可能引起选择偏见，影响 O 的实际分布，与理想测试分布不符。AMID通过适应性偏差校正来纠正这种偏差，确保在复杂环境下优化推荐准确性。

Single-domain Sequential Recommendation Methods

在研究中，我们旨在构建一个跨领域的通用模型，提升单领域序列推荐的性能。为此，我们首先研究了单域推荐的基础机制。嵌入层通过截断或填充将序列 S 映射到维度为 d 的嵌入空间，同时使用可学习的动态位置嵌入来捕捉时间顺序信息。我们使用序列信息编码器，如GRU4Rec、SASRec和BERT4Rec，通过 \mathcal{F} 函数来提取序列中的短期和长期项目关系。这样做的目的是将这些模型的序列特征整合到一个全面的CDSR模型中，以减少领域间数据选择偏见的影响，确保在开放世界环境下的在线推荐有效性。

Methodology



Multi-interest Information Module

在这个模块中，我们构建多兴趣信息，将单一的SDSR模型转变为CDSR模型。步骤包括兴趣群组构建和信息传播。以两个领域为例，我们通过比较用户序列来生成群组标识。对于用户 u_i 和 u_j ，

$$\text{group_flag}_{ij} = \text{similarity}(S_{u_i}^{Z_1}, S_{u_j}^{Z_2})$$

这里的**similarity**是一个衡量序列相似性的函数，它通过比较序列中的点来判断用户是否存在相似的兴趣偏好，以决定是否将他们合并到同一兴趣群组。这个过程是通用的，适用于多个领域，旨在通过群组共享信息来减少领域间数据偏见，提高跨领域推荐的有效性。

$$\mathbf{a}'_{ij} = \max[(S_{u_i}^{Z_1} \mathbf{W}_1)(S_{u_j}^{Z_2} \mathbf{W}_2)^\top]$$

在这个方法中，我们利用矩阵 \mathbf{W}_1 和 \mathbf{W}_2 对用户在 Z_1 和 Z_2 领域的序列进行变换，通过计算**内积**来量化相似性。具体来说，我们通过：

$$\text{similarity} = (\mathbf{W}_1 \cdot S_{u_i}^{Z_1})^\top \cdot (\mathbf{W}_2 \cdot S_{u_j}^{Z_2})$$

找到用户 u_i 和 u_j 之间的最相似程度。若该值超过阈值 \mathbf{a}_{ij} ，则定义为 $\mathbf{a}_{ij} = 1$ ，表示他们属于同一兴趣群组。通过这种方式，我们在多领域场景下通过比较和匹配用户序列来构建兴趣群组，从而将SDSR模型转换为CDSR模型，减少领域间数据偏见，提升跨领域推荐的准确性。

我们从用户 $u_i^{Z_1}$ 和 $u_j^{Z_2}$ 分别提取各自的序列特征，通过应用共享的**变换矩阵** \mathbf{W}_1 和 \mathbf{W}_2 ，计算它们在 d 维空间中的相关性，即：

$$\mathbf{m}_{u_i \leftarrow u_j^{Z_2}} = (\mathbf{W}_1 \cdot S_{u_i}^{Z_1}) \circ (\mathbf{W}_2 \cdot S_{u_j}^{Z_2})$$

这里，Hadamard product保留了序列中对应位置元素的乘积，形成跨域特征交互。这样产生的信息作为用户间的跨域桥梁，被整合进各自的嵌入中，以增强CDSR模型对不同领域知识的融合和利用。

$$\mathbf{m}_{u_i \leftarrow u_j^{Z_2}} = \mathbf{a}_{ij} \cdot (S_{u_j}^{Z_2} \mathbf{W}_{ip})$$

信息传播阶段在群组内进行。首先，对于领域 Z_1 的用户 $u_i^{Z_1}$ 和领域 Z_2 的用户 $u_j^{Z_2}$ ，通过学习的矩阵 \mathbf{W}_{ip} ，将跨域信息从 Z_2 传递到 Z_1 。如果用户在多个领域有行为，跨域行为 $\mathbf{m}_{u_i \leftarrow u_j^{Z_2}}$

也被纳入考虑。针对目标用户 $u_i^{Z_1}$ ，我们收集所有领域的信息，将这些信息按行为时间与领域维度堆叠成 $\mathbf{m}'_{u_i} \in \mathbb{R}^{T \times d \times N}$

然后，使用矩阵 \mathbf{W}_C 和 \mathbf{W}_F 融合这些跨域信息。对于其他领域用户，传播过程亦遵循类似步骤，但使用**Squeeze**函数来减小矩阵维度。最终，生成增强的序列表示 $S_{u_i}^{*Z}$ ，整合了初始序列和跨域信息。

$$S_{u_i}^{*Z} = \text{Concat}(S_{u_i}^Z, \text{Squeeze}(\mathbf{m}'_{u_i} \mathbf{W}_C)) \mathbf{W}_F$$

Prediction Layer

预测层用于预测用户 u_i 对目标物品 v_k 的偏好，公式为： $\hat{p}_{i,k} = f(S_{u_i}^{*Z}, \mathbf{V}_k; \theta_p)$

这里 $\hat{p}_{i,k}$ 是预测得分 f 是ReLU或softmax这样的**非线性激活函数**。用户的行为信息和物品特征通过 $S_{u_i}^{*Z}$ （融合了跨域信息的用户表示）以及物品的**特征向量** \mathbf{V}_k 输入。参数集 θ_p 通过训练学习用户行为模式和物品特性，以更精确地预测用户对未接触过物品的兴趣，从而优化了CDSR模型的跨领域推荐效果。

CDSR模型使用**多层感知器**（MLPs）结构，它将合并的用户序列特征（ $\mathbf{E}_{u_i}^{*Z}$ ）和物品嵌入（ \mathbf{V}_k ）作为输入。MLPs通过堆叠多层全连接神经单元，对输入进行加权处理，每层使用sigmoid激活函数进行非线性转换。 \oplus 操作将序列和嵌入拼接。具体计算步骤如下：首先，输入通过前一层次的**权重矩阵** $\mathbf{W}^{(l)}$ 和**偏置项** $\mathbf{b}^{(l)}$ 进行计算，得到输出 $\mathbf{x}_{i,k}^{(l+1)}$ 。然后，对所有时间步长的输出求平均，通过Mean函数得到：

$$\mathbf{x}_{i,k}^{(L)} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_{i,k}^{(L-1)}(t)$$

Doubly Robust Estimator for CDSR

我们提出了一种创新的双稳健估计器，适用于跨域序列环境，扩展了传统方法。设 $\hat{\mathbf{R}}^Z$ 为预测矩阵，真实评分矩阵为 \mathbf{R}^Z 。评估指标包括预测误差 \mathcal{P} 和双稳健估计器

$$\mathcal{E}_{\text{DR}}^*$$

$$\mathcal{P} = \|\hat{\mathbf{R}}^Z - \mathbf{R}^Z\|_F^2 \quad (\text{预测误差})$$

$$\mathcal{E}_{\text{DR}}^* = \mathbf{R}^Z (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{R}}^Z + (\hat{\mathbf{R}}^Z - \mathbf{R}^Z) \cdot (\mathbf{I} - \mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \mathbf{R}^Z$$

其中 \mathbf{X} 是用户行为序列矩阵 \mathbf{x}_{ui}^Z 表示用户在不同时间对所有物品的实际观测。双稳健估计器通过同时考虑直接基于预测的估计和潜在观测项的修正，提供更稳定的估计，减少了预测误差的影响。

$$\mathcal{P} = \frac{1}{|Z|} \sum_{Z \in \mathcal{Z}} \frac{1}{|\mathcal{D}^Z|} \sum_{u,v \in \mathcal{D}^Z} e_{u,v}^Z$$

$$\mathcal{E}_{\text{DR}}^* = \frac{1}{|Z|} \sum_{Z \in \mathcal{Z}} \frac{1}{|\mathcal{D}^Z|} \sum_{u,v \in \mathcal{D}^Z} \left(\hat{e}_{u,v}^Z + \frac{\alpha_{u,v}^Z \delta_{u,v}^Z}{\hat{p}_{u,v}^Z} \right)$$

$$\text{Bias}_{\text{DR}} = \mathbf{E}[\hat{\mathbf{E}}^Z] - \mathbf{E}[(\mathbf{1} - \hat{\mathbf{P}}^Z) \mathbf{E}[\mathbf{R}^Z | \mathbf{S}_u^{*Z}, \mathbf{V}^Z]] + \mathbf{E}[\hat{\mathbf{P}}^Z (\mathbf{R}^Z - \hat{\mathbf{R}}^Z)]$$

其中 $\mathbf{E}[\cdot]$ 代表期望值 $\mathbf{1}$ 是全一维单位向量 $\mathbf{E}[\mathbf{R}^Z | \mathbf{S}_u^{*Z}, \mathbf{V}^Z]$

是基于用户行为和物品特征时真实评分矩阵的期望值。DR通过整合imputation模型 ($\hat{\mathbf{E}}^Z$) 和潜在观测的修正 (通过

$$(\mathbf{1} - \hat{\mathbf{P}}^Z)(\mathbf{R}^Z - \hat{\mathbf{R}}^Z)$$

，降低了因预测误差产生的偏差，提高了估计的稳健性。这种多任务的学习策略使得DR估计器更具抗干扰能力。

$$\text{Bias}(\mathcal{E}_{\text{DR}}^*) = \frac{1}{|Z|} \sum_{Z \in \mathcal{Z}} \left[\frac{1}{|\mathcal{D}^Z|} \sum_{u,v \in \mathcal{D}^Z} \Delta_{u,v}^Z \delta_{u,v}^Z \right]$$

在这个上下文中 $\delta_{u,v}^Z$ 和 $\Delta_{u,v}^Z$ 分别是用于评估预测误差的两个指标。它们定义如下：

$$\delta_{u,v}^Z = g_{\phi^Z}(\text{Mean}(\mathbf{S}_u^{*Z}) || \mathbf{v}^Z) - (r_{u,v}^Z - \hat{r}_{u,v}^Z)$$

$$\Delta_{u,v}^Z = g_{\psi^Z}(\text{Mean}(\mathbf{S}_u^{*Z}) || \mathbf{v}^Z) - \hat{p}_{u,v}$$

其中 $\text{Mean}(\mathbf{S}_u^{*Z})$ 是通过用户行为序列增强的表示 \mathbf{v}^Z 是物品的嵌入 $\hat{r}_{u,v}^Z$ 是对用户对物品潜在观察的估计，实际评分是 $r_{u,v}^Z$ $\hat{p}_{u,v}$ 是用户对物品的倾向性估计。 g_{ϕ^Z} 和 g_{ψ^Z} 是用于imputation（缺失值填充）和propensity model（倾向性估计）的函数。这两个指标用于计算在实际数据上的预测误差，然后在DR估计中用来校正偏差，以提高估计的准确性。

$$\Delta_{u,v}^Z = \frac{\hat{p}_{u,v}^Z - p_{u,v}^Z}{\hat{p}_{u,v}^Z}, \quad \delta_{u,v}^Z = e_{u,v}^Z - \hat{e}_{u,v}^Z$$

定理4.1表述了双重稳健性：当所有用户-物品的填充误差 $\hat{\mathbf{E}}^Z$ 和倾向性估计 $\hat{\mathbf{P}}^Z$ 都精确时，CDSR的DR估计量（差分隐私保护的差分回归估计）是无偏的，即不会偏离真实期望值。推论4.2给出了DR估计器的一个尾部界限。假设给定的误差存在，如果对于任意预测矩阵 $\hat{\mathbf{R}}^Z$ ，在CDSR任务中，DR估计的偏差在90% $(1-\eta)$ 的概率下，其绝对误差不超过 t ，则有：

$$\Pr(|\mathcal{E}_{\text{DR}}^* - \mathbf{E}[\mathcal{E}_{\text{DR}}^*]| > t) \leq \eta, \quad \text{对任意 } t > 0$$

$$|\mathcal{E}_{DR}^* - \mathbb{E}_O[\mathcal{E}_{DR}^*]| \leq \sqrt{\frac{\log(\frac{2}{\eta})}{2|Z|(\sum_{z \in Z} |\mathcal{D}^z|)^2} \sum_{z \in Z} \left[\frac{1}{|\mathcal{D}^z|} \sum_{u,v \in \mathcal{D}^z} \left(\frac{\delta_{u,v}^z}{\hat{p}_{u,v}^z} \right)^2 \right]}$$

推论4.2进一步指出，相比于使用IPS估计器的IPSCDR，当填充误差满足 $\hat{e}_{u,v}^z \leq 2e_{u,v}^z$

DR估计器的尾部界限更小。这意味着，即使仅考虑填充误差的严格约束，DR估计器在确保稳健性方面仍优于IPS，因为它能有效地减少预测误差带来的偏差。详细的证明可在附录中查阅。

Joint learning

我们采用交替训练策略进行联合学习，首先在已知数据上，通过优化针对提出的混合损失函数⁺，同时训练补全模型和预测模型⁺。

$$\begin{aligned} \mathcal{L}_e(\theta, \phi, \psi) = & \frac{1}{|Z|} \sum_{z \in Z} \left(\frac{1}{|\mathcal{O}^z|} \sum_{u,v \in \mathcal{O}^z} e_{u,v} + \lambda_1 \sum_{u,v \in \mathcal{O}^z} \frac{(\hat{e}_{u,v} - e_{u,v})^2}{\hat{p}_{u,v}} \right) \\ & + \lambda_2 \|\theta\|_F^2 + \lambda_3 \|\phi\|_F^2 + \lambda_4 \|\psi\|_F^2 \end{aligned}$$

我们利用Frobenius范数⁺ $\|\cdot\|_F$ 度量矩阵范数⁺，通过超参数⁺ $\lambda_{1,2,3,4}$ 调整正则化⁺与多任务损失的相对重要性。在对已观测数据进行训练后，对预测模型 θ 在 \mathcal{D}^Z 上进行进一步训练，目的是减少偏差。

$$\begin{aligned} \mathcal{L}_r(\theta, \phi, \psi) = & \frac{1}{|Z|} \sum_{z \in Z} \left[\frac{1}{|\mathcal{D}^z|} \sum_{u,v \in \mathcal{D}^z} \left(\hat{e}_{u,v} + \frac{o_{u,v}(e_{u,v} - \hat{e}_{u,v})}{\hat{p}_{u,v}} \right) \right] \\ & + \lambda_5 \|\theta\|_F^2 \end{aligned}$$

对于集外数据集 $\mathcal{D}^Z - \mathcal{O}^Z$

我们无法直接获取标签 $y_{u,v}$ ，这表明我们只能在已观察数据子集 \mathcal{O}^Z 上进行学习。为了控制偏差，我们引入超参数 λ_5 来平衡正则化项。整个学习过程的具体细节在附录中有详细说明。

Experiments

Experimental Setup

实验中，我们沿用前人做法，利用亚马逊数据集（链接提供），选取'Cloth-Sport'和'Phone-Elec'两个类别，构建两个不同的任务。为模拟在线平台的偏见，我们调整不重叠用户比例 κ_u ，取值区间为25%-75%。高比例意味着更接近无偏差环境。评估采用标准的NDCG@10和精确率（HR@10），这两种指标在CDSR领域常用。为确保公平，我们随机选取999个非互动负样本与1个正样本进行排名测试。每种设置重复五次以计算平均结果和方差。详细实验信息请参阅附录。

Performance Comparisons

我们对比了五种方法：(1) 单领域模型⁺BERT4Rec、GRU4Rec和SASRec；(2) 传统跨领域模型STAR、MAMDR和SSCDR；(3) 跨领域模型Pi-Net、DASL和C²DSR；以及(4) 偏见纠正模型DCRec和IPSCDR。AMID方法因其兼顾重叠和非重叠用户，以及与SDSR兼容性，展现出最通用性和普遍性。在偏差校正方面，AMID有效处理各类偏差，尤其是选择偏差，且其提出的DR估计器在CDSR中具有较低的方差，对比IPSCDR中的IPS，这个优势明显。

	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10	NLA_Avg10	FIRGP10
BERT4Rec [48]	1.42±0.13	2.96±0.15	2.28±0.19	4.43±0.46	2.96±0.25	5.60±0.34	4.19±0.18	7.77±0.24	6.13±0.16	10.87±0.24	6.20±0.17	11.07±0.47	8.24±0.22	13.05±0.33	10.58±0.07	17.28±0.05
GRU4Rec [19]	2.03±0.23	4.19±0.35	3.23±0.15	4.18±0.37	3.63±0.22	7.35±0.47	4.85±0.16	9.43±0.20	6.99±0.23	12.98±0.39	6.98±0.18	12.90±0.38	10.11±0.12	16.77±0.23	11.41±0.08	19.00±0.12
SASRec [24]	2.00±0.13	4.26±0.23	3.34±0.18	6.66±0.38	3.69±0.32	7.46±0.55	4.96±0.30	9.69±0.38	7.19±0.19	13.32±0.30	7.14±0.26	13.34±0.34	10.56±0.18	17.77±0.21	11.64±0.04	19.67±0.18
STAR [45]	1.97±0.27	4.19±0.39	3.37±0.15	6.54±0.28	3.60±0.29	7.10±0.51	4.77±0.18	9.29±0.15	6.96±0.24	13.30±0.45	7.13±0.18	13.71±0.51	9.72±0.17	16.58±0.26	11.18±0.13	19.00±0.29
MAMDR [34]	2.11±0.09	4.25±0.17	3.44±0.15	6.60±0.15	3.53±0.44	7.21±0.24	4.84±0.14	9.33±0.42	7.01±0.22	13.38±0.42	7.19±0.19	13.81±0.20	9.79±0.31	16.47±0.39	11.24±0.47	19.08±0.25
SSCDR [23]	2.02±0.18	4.21±0.36	3.42±0.24	6.57±0.17	3.45±0.24	7.14±0.40	4.81±0.29	9.31±0.57	6.99±0.39	13.35±0.45	7.21±0.35	13.83±0.30	9.75±0.27	16.45±0.27	11.27±0.31	19.13±0.33
Pi-Net [36]	1.84±0.26	3.82±0.38	2.77±0.20	5.56±0.32	3.37±0.17	7.05±0.22	4.56±0.25	8.81±0.34	7.02±0.26	12.68±0.50	7.11±0.26	12.95±0.36	10.00±0.11	16.33±0.28	11.64±0.06	19.34±0.26
DASL [27]	2.11±0.29	4.56±0.43	3.32±0.17	6.62±0.29	3.96±0.16	8.11±0.37	4.81±0.39	9.71±0.51	7.10±0.15	13.29±0.17	7.22±0.13	13.22±0.17	10.35±0.33	17.15±0.57	11.72±0.07	19.52±0.18
C²DR [25]	2.27±0.18	4.73±0.35	3.31±0.07	6.60±0.24	3.74±0.26	7.99±0.42	5.18±0.10	10.31±0.07	7.54±0.15	14.04±0.20	7.30±0.19	13.96±0.44	10.71±0.13	17.98±0.10	11.74±0.03	20.04±0.21
DCRC [60]	2.22±0.14	4.37±0.12	3.40±0.08	6.51±0.12	3.85±0.14	7.72±0.27	5.21±0.15	10.08±0.26	7.18±0.08	13.58±0.13	7.15±0.07	13.32±0.25	10.48±0.14	17.84±0.22	11.79±0.08	20.01±0.09
BERT4Rec [48] + CaseQ [65]	2.24±0.16	4.41±0.13	3.45±0.07	6.48±0.13	3.82±0.15	7.69±0.29	5.28±0.16	10.04±0.27	7.25±0.07	13.52±0.14	7.19±0.06	13.33±0.26	10.56±0.13	17.72±0.23	11.88±0.07	19.98±0.10
GRU4Rec [19] + CaseQ [65]	2.20±0.15	4.50±0.25	3.58±0.09	6.57±0.19	3.98±0.18	7.76±0.46	5.30±0.14	10.08±0.18	7.28±0.18	13.34±0.29	7.38±0.17	13.52±0.28	10.59±0.15	17.88±0.23	11.98±0.07	20.04±0.14
SASRec [24] + CaseQ [65]	2.23±0.23	4.46±0.12	3.51±0.07	6.56±0.26	3.87±0.14	7.75±0.46	5.32±0.16	10.21±0.36	7.31±0.12	13.47±0.23	7.37±0.11	13.49±0.15	10.60±0.12	17.72±0.16	11.93±0.06	19.98±0.14
BERT4Rec [48] + IPSCDR [29]	1.95±0.11	3.96±0.20	2.82±0.31	5.55±0.73	3.65±0.22	7.06±0.33	4.95±0.29	9.46±0.21	7.29±0.19	13.20±0.32	7.47±0.33	13.46±0.07	9.71±0.24	15.75±0.30	11.94±0.15	19.66±0.20
GRU4Rec [19] + IPSCDR [29]	2.53±0.13	4.93±0.44	3.79±0.18	7.11±0.42	4.24±0.28	8.50±0.36	5.64±0.16	10.60±0.20	7.58±0.22	13.90±0.49	7.84±0.26	14.16±0.54	10.69±0.07	17.78±0.12	12.18±0.10	20.33±0.17
SASRec [24] + IPSCDR [29]	2.48±0.10	4.92±0.22	3.67±0.08	6.82±0.22	4.10±0.26	8.13±0.46	5.55±0.15	10.39±0.26	7.61±0.13	14.05±0.15	7.95±0.14	14.27±0.19	10.69±0.12	18.03±0.12	12.26±0.12	20.36±0.32
BERT4Rec [48] + AMID	1.80±0.23	3.68±0.41	2.24±0.13	4.48±0.20	2.83±0.18	5.45±0.31	4.11±0.31	7.79±0.47	6.83±0.13	11.81±0.25	6.84±0.14	12.11±0.34	8.41±0.13	13.38±0.16	11.23±0.17	18.39±0.34
GRU4Rec [19] + MM	2.58±0.07	4.88±0.09	3.63±0.13	6.61±0.30	4.47±0.30	8.50±0.40	5.62±0.18	10.58±0.25	7.88±0.23	14.31±0.34	7.75±0.32	14.39±0.30	11.66±0.17	18.18±0.25	12.21±0.10	20.32±0.17
SASRec [24] + MM	2.67±0.24	5.11±0.36	3.88±0.24	7.03±0.45	4.48±0.25	8.74±0.30	5.66±0.18	10.58±0.26	7.99±0.17	14.77±0.13	7.84±0.18	14.36±0.33	11.31±0.11	18.69±0.14	12.32±0.14	20.67±0.30
BERT4Rec [48] + AMID	2.99±0.07	5.70±0.13	3.79±0.15	7.09±0.32	4.73±0.29	9.30±0.41	5.70±0.12	10.63±0.36	7.95±0.23	14.49±0.36	8.15±0.54	14.18±0.28	11.74±0.07	18.52±0.21	12.18±0.10	20.10±0.31
GRU4Rec [19] + AMID	3.10±0.17	5.95±0.16	3.94±0.15	7.30±0.30	4.89±0.10	9.42±0.28	5.90±0.17	11.01±0.17	8.33±0.09*	15.49±0.50*	8.34±0.27*	14.29±0.29*	11.74±0.07	18.52±0.21	12.18±0.10	20.10±0.31
SASRec [24] + AMID	3.20±0.22*	6.14±0.33*	4.19±0.10*	7.62±0.20*	4.97±0.15*	9.48±0.22*	5.96±0.14*	11.04±0.26*	8.20±0.10	14.99±0.16	8.32±0.25*	14.96±0.15	11.71±0.23	19.28±0.23	12.52±0.13	20.79±0.20
Improvement(%)	26.48	34.54	10.55	7.17	17.22	11.53	5.67	6.15	9.46	10.25	4.91	7.15	6.70	8.35	2.20	2.41

定量评估：表中呈现了在两种不同程度选择偏差的CDSR任务中，各模型的量化比较。 K_u 数值越大，偏差越轻微。粗体标记和下划线分别表示最优和次优结果。随 K_u 增加，多数情况下，模型性能普遍提升，因为更倾斜的环境增加了挑战。我们的观察如下模型效率分析：所有模型在相同配置的NVIDIA GeForce A100 GPU（80GB内存，Intel Core i7-8700K CPU，64GB内存）上进行训练和测试。

Table 3: Statistics on the Amazon datasets.

Dataset		Users	Items	Ratings	#Overlap	Avg.length	Density
Amazon	Cloth	27,519	9,481	161,010	16,337	4.39	0.06%
	Sport	107,984	40,460	851,553		7.58	0.02%
Amazon	Phone	41,829	17,943	194,121	7,857	4.53	0.03%
	Elec	27,328	12,655	170,426		6.19	0.05%
MYbank	Loan	39,557,003	61,934	227,079,281	29,476,198	1.82	0.01%
	Fund	48,439,382	13,927	133,836,385		2.57	0.02%
MYbank	Loan	39,557,003	61,934	227,079,281	37,821,145	1.82	0.01%
	Account	92,692,975	21,599	278,948,331		2.31	0.01%

尽管C²DSR、SASRec+CaseQ、SASRec+IPSCDR和SASRec+AMID的参数量接近，分别为0.276M、0.210M、0.192M和0.193M，但AMID在处理单个批次样本时，训练和测试的时间效率明显优于其他模型，分别为0.111s/0.032s，远低于C²DSR的0.130s/0.049s和SASRec+IPSCDR的0.143s/0.045s。这表明AMID在保证性能的同时，有效提高了计算资源的利用率，展现出在开放世界CDSR任务中的高效性。

Online A/B Test

用户接触服务的次数（#exposure，曝光率提升9.65%）、用户点击服务的次数（#clk，点击率提升5.69%）以及转化率（CVR，提升1.32%）。为期十四天的测试结果显示，我们的方法在所有领域都带来了显著改善。

Table 6: Online A/B testing results from 9.15 to 9.28, 2023

	# exposure	# clk	CVR
Loan Domain	+9.89%	+5.27%	+1.42%
Fund Domain	+7.32%	+4.94%	+0.98%
Account Domain	+11.73%	+6.85%	+1.57%

原文《Rethinking Cross-Domain Sequential Recommendation under Open-World Assumptions》

编辑于 2024-05-24 15:05 · IP 属地北京

序列推荐 推荐系统 开放世界