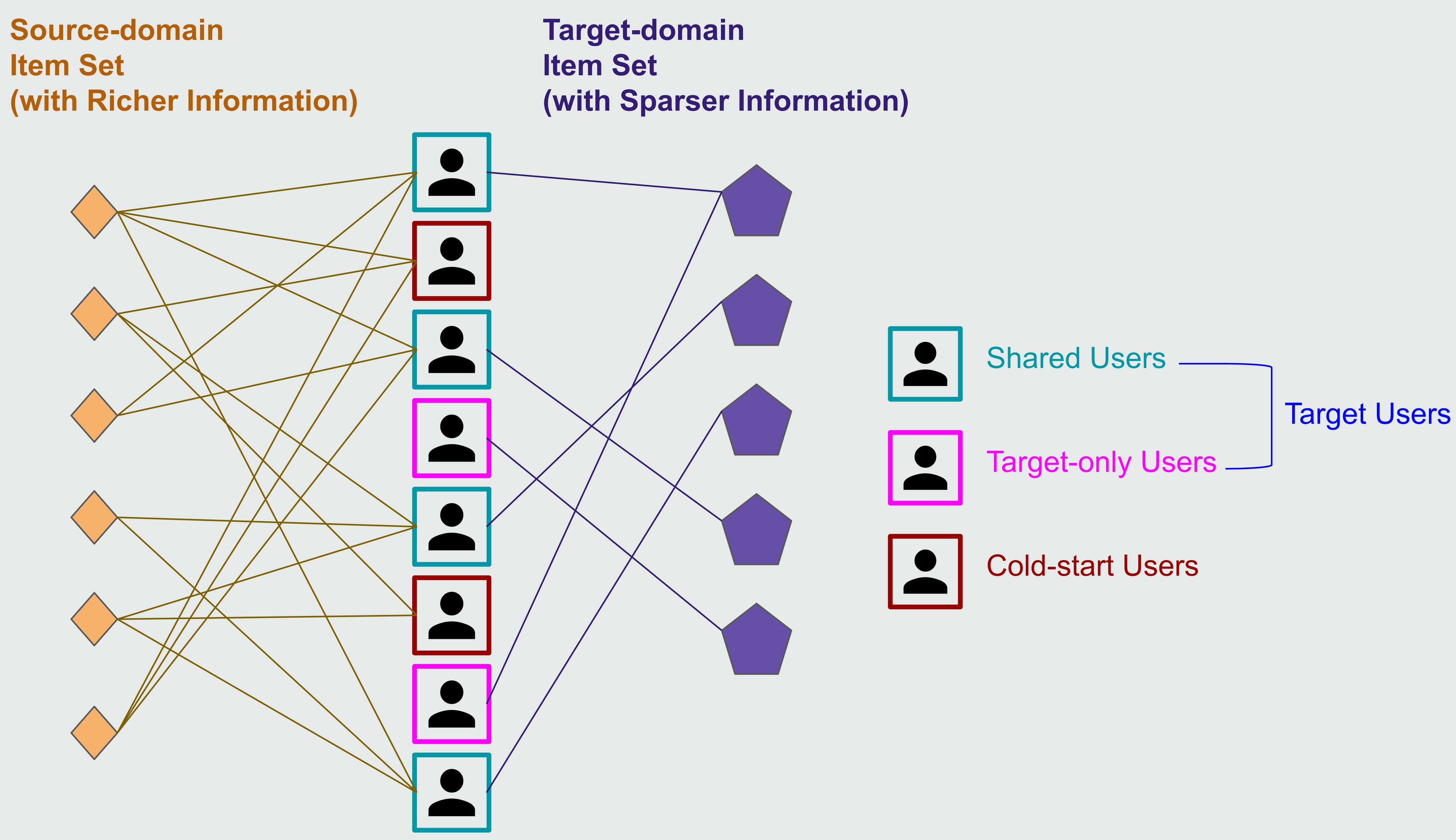
CPR: Cross-domain Preference Ranking with User Transformation

Yu-Ting Huang[1,2] Hsien-Hao Chen[3] Tung-Lin Wu[2] Chia-Yu Yeh[2] Jing-Kai Lou[4] Ming-Feng Tsai[3] Chuan-Ju Wang[2]

[1]National Taiwan University [2]Academia Sinica [3]National Chengchi University [4]KKStream Limited

Abstract

- Data sparsity is a well-known challenge in recommender systems. One way to alleviate this problem is to leverage knowledge from relevant domains.
- Although several studies leverage side information (e.g., user reviews) for cross-domain recommendation, side information is not always available or easy to obtain in practice.
- To this end, we propose cross-domain preference ranking (CPR) with a simple yet effective user transformation that leverages *only* user interactions with items in the source and target domains to transform the user representation.
- Given the proposed user transformation, CPR not only successfully enhances recommendation performance for users having interactions with target-domain items but also yields superior performance for cold-start users in comparison with state-of-the-art cross-domain recommendation approaches.



Problem Definition

We consider the recommendation scenario involving two domains with disjoint item sets, namely, a source-domain item set and a target-domain item set(denoted as I^{S} and I^{T}, respectively); there exists a set of users having interactions with items from both domains, namely *shared users*.

Formally, we denote the set of users having interactions with items in I^{S} (I^{T}) as U^{S} (U^{T}, respectively) and the shared users as $U^{\text{shared}} = U^{\mathrm{S}} \cap U^{\mathrm{T}}$ and $U^{\text{shared}} \neq \emptyset$.

Let $I = I^{\mathrm{S}} \cup I^{\mathrm{T}}$ and $U = U^{\mathrm{S}} \cup U^{\mathrm{T}}$. The goal of the proposed CPR approach is to learn the representation matrix $\Theta \in \mathbb{R}^{(|U|+|I|)\times d}$ mapping each user and item to a d-dimensional embedding vector.

Proposed CPR Approach

Given a user u, let I_u^{S} (I_u^{T}) denote the set of items in the source domain (target domain, respectively) that u has interacted with. To transfer knowledge from the source domain into the target domain, we bridge the non-overlapped I^{S} and I^{T} with the following user representation transformation: for each user $u \in U$, we have

$$\Theta_u = \Theta_u^{\text{pseudo}} + \vec{a}_{I_u^{\mathrm{S}}} + \vec{a}_{I_u^{\mathrm{T}}},$$

in which Θ_u^{pseudo} denotes a learnable pseudo user representation for user u, $\vec{a}_{I_u^{\mathrm{S}}} = 1/|I_u^{\mathrm{S}}| \sum_{i \in I_u^{\mathrm{S}}} \Theta_i$, and $\vec{a}_{I_u^{\mathrm{T}}} = 1/|I_u^{\mathrm{T}}| \sum_{i \in I_u^{\mathrm{T}}} \Theta_i$.

With the above transformation , we formulate the maximum posterior estimator to derive our optimization criterion for CPR as

$$\text{CPR-OPT} := \sum_{u \in U^{\mathrm{T}}} \sum_{\substack{t^+ \in I_u^{\mathrm{T}} \\ t^- \in I^{\mathrm{T}} \setminus I_u^{\mathrm{T}}}} \ln \sigma\left(\langle \Theta_u, (\Theta_{t^+} - \Theta_{t^-}) \rangle\right) - \lambda ||\Theta||^2,$$

where $\sigma(\cdot)$ denotes the sigmoid function, $\langle \cdot, \cdot \rangle$ denotes the inner product for two vectors, and λ is a regularization parameter.

- For more details, please refer to: `https://link.springer.com/chapter/10.1007/978-3-031-28238-6_35`

Result

	HK-CSJ		MT-B		SPO-CSJ	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
BPR	0.4403	0.3080	0.5254	0.3324	0.4289	0.2905
BPR+	0.3674	0.2381	0.5203	0.3316	0.4006	0.2660
LightGCN	0.5117	0.3945	0.8454	0.6736	0.5077	0.3824
LightGCN+	†0.5377	†0.4070	†0.8594	†0.6820	0.5217	0.3877
EMCDR	0.4106	0.2775	0.5166	0.3266	0.4266	0.2888
Bi-TGCF	0.5369	0.3939	0.8391	0.6424	†0.5520	†0.4020
CPR	***0.5677**	***0.4290**	***0.8954**	***0.7145**	**0.5534**	**0.4183**
Improv.	5.58%	5.42%	4.19%	4.76%	0.26%	4.05%

Table 1. Test users from target users

	HK-CSJ		MT-B		SPO-CSJ	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
BPR	0.2837	0.1750	0.1874	0.1143	0.2249	0.1330
BPR+	0.2560	0.1405	0.1874	0.1140	0.2186	0.1208
LightGCN	0.3520	0.2450	†0.4263	†0.3216	0.3803	0.2640
LightGCN+	†0.3714	†0.2508	0.4160	0.3128	0.3674	0.2566
EMCDR	0.2566	0.1434	0.2089	0.1250	0.1680	0.0861
Bi-TGCF	0.3583	0.2368	0.4174	0.2925	†0.3900	†0.2662
CPR	***0.3929**	***0.2729**	***0.4594**	***0.3441**	***0.4154**	***0.2929**
Improv.	5.77%	8.81%	7.77%	7.00%	6.52%	10.01%

Table 2. Test users from shared users

	HK-CSJ		MT-B		SPO-CSJ	
	HR@10	NDCG@10	HR@10	NDCG@10	HR@10	NDCG@10
BPR+	0.2417	0.1327	0.1351	0.0810	0.1806	0.0942
LightGCN+	0.1380	0.0748	0.0580	0.0287	0.1386	0.0833
EMCDR	†0.2514	†0.1407	†**0.2034**	†**0.1203**	0.1466	0.0762
Bi-TGCF	0.2477	0.1370	0.1211	0.0686	†0.2569	†0.1548
CPR	***0.3160**	***0.1899**	*0.1760	*0.1014	***0.3371**	***0.2100**
Improv.	25.68%	34.90%	-13.48%	-15.68%	31.26%	35.62%

Table 3. Test users from cold-start users

In the tables, the best performance is in boldface; '†' indicates the best performing method among all the baselines; '*' and 'Improv. (%)' denote statistical significance at $p < 0.05$ with a paired t-test and the percentage improvement of our model, respectively, with respect to the best performing baseline.