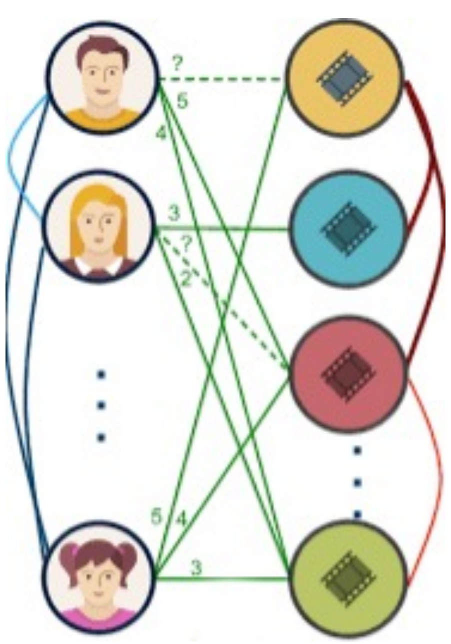
IPR: Interaction-level Preference Ranking for Explicit Feedback

Shih-Yang Liu[1], Hsien Hao Chen[2], Chih-Ming Chen[2], Ming-Feng Tsai[2], Chuan-Ju Wang[1]
[1]Academia Sinica, [2]National Chengchi University Taipei, Taiwan

Introduction

Explicit feedback—user input regarding their interest in an item—is the most helpful information for recommendation as it comes directly from the user and shows their direct interest in the item. Most approaches either treat the recommendation given such feedback as a typical regression problem or regard such data as implicit and then directly adopt approaches for implicit feedback; both methods, however, tend to yield unsatisfactory performance in top-k recommendation. In this paper, we propose interaction-level preference ranking (IPR), a novel pairwise ranking embedding learning approach to better utilize explicit feedback for recommendation. Experiments conducted on three real-world datasets show that IPR yields the best results compared to six strong baselines.



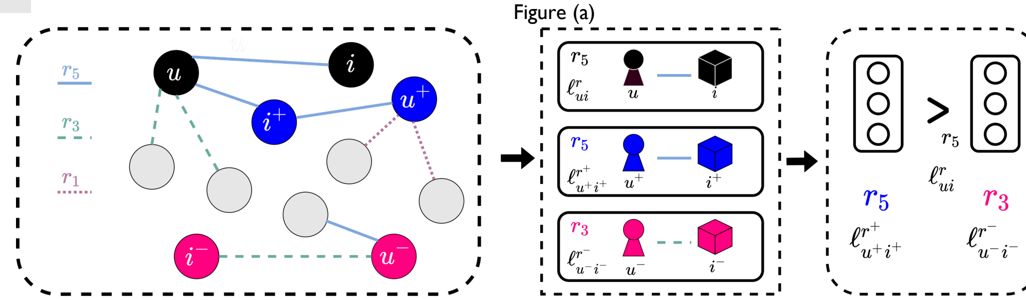
Some key terms:
collaborative filtering; matrix factorization; explicit feedback; top-k recommendation; high-order graph information

Dataset: Movielens-1M, Amazon Review Datasets
Code: https://github.com/seantheplug/SIGIR_2022_IPR

Methodology

To better exploit the information encoded in different magnitudes of user preferences (explicit feedback) for recommendation, we propose interaction-level preference ranking (IPR), a unified and interaction-level embedding learning framework for explicit feedback. In particular, IPR changes the main idea of many pairwise ranking recommendation algorithms from node-level modeling to interaction-level modeling and clusters interactions with similar (or same) ratings in a self-supervised manner for explicit feedback.
To construct the framework, we leverage Bayesian preference ranking (BPR), a prevalent concept in the recommendation literature for implicit feedback. The basic training unit for a typical BPR-based algorithm is the node triplet (u, i, j) and relation $i >_u j$ denotes that user u prefers item i over item j. Using a similar concept, we change the picture by altering the basic unit—the node triplet—to an interaction triplet. Let L be the set containing all user-item interactions, where each element $l^r_{ui} \in L$ denotes a user-item interaction in which user u interacts with item i with r as the user rating.

Figure (a)



Then the basic trainig unit can be formulated as the middle part of Fig (a), where the blue interaction is considered "more similar" to the target interactionn while the pink interaction is considered "more irrelevant". We propose a sampling stratgy to construct the basic trainig unit. To elaborate, we take Fig (a) as an example. Given an interaction between user u and i with a specific rating r_5, we first sample an item i^+ that user u has interacted with and to which she has assigned an identical rating r_5, according to which we further sample a user u^+ who has interacted with item i^+ and given the same rating r_5 to construct a positive interaction (blue interaction) regarding to the black target interaction. For the negative interaction, we simply set it to be any interaction but the possible positive interactions to the target interaction. An interaction is a combination of user embedding and an item embedding, then we optimize the embeddings (matrix) with maximum posterior estimator to maximize the log likelihood of the positive interaction regarding the target interaction. After training, dot product between a user embedding and an item embedding is calculated as the score.

Results

	Movielens-1M						AMZ H-PC						AMZ BT					
	Recall			nDCG			Recall			nDCG			Recall			nDCG		
	@1	@3	@10	@1	@3	@10	@1	@3	@10	@1	@3	@10	@1	@3	@10	@1	@3	@10
MF	0.0097	0.0287	0.0894	0.2241	0.2195	0.2169	0.0088	0.0178	0.0382	0.0151	0.0175	0.0254	0.0090	0.0242	0.0641	0.0378	0.0354	0.0499
MF-BPR	0.0133	0.0376	0.1061	0.2736	0.2676	0.2547	0.0117	0.0245	0.0477	0.0210	0.0240	0.0328	0.0110	0.0281	0.0711	0.0433	0.0405	0.0561
LightGCN	0.0209	0.0536	0.1316	0.4463	0.4100	0.3524	†0.0146	†0.0316	†0.0691	†0.0294	†0.0320	†0.0459	†0.0119	†0.0327	†0.0884	0.0496	†0.0458	†0.0669
DeepICF	†0.0221	†0.0562	0.1362	†0.4625	†0.4209	†0.3634	0.0052	0.0142	0.0389	0.0141	0.0153	0.0240	0.0112	0.0286	0.0726	†0.0511	0.0452	0.0590
CDAE	0.0204	0.0543	†0.1431	0.4241	0.3990	0.3581	0.0073	0.0186	0.0435	0.0131	0.0170	0.0269	0.0112	0.0267	0.0705	0.0457	0.0398	0.0552
MCBPR	0.0116	0.0288	0.0672	0.2793	0.2451	0.2015	0.0070	0.0146	0.0280	0.0137	0.0149	0.0197	0.0068	0.0165	0.0453	0.0308	0.0267	0.0359
IPR	**0.0270**	**0.0666**	**0.1564**	**0.5124**	**0.4610**	**0.3943**	**0.0163**	**0.0359**	**0.0713**	**0.0309**	**0.0353**	**0.0486**	**0.0132**	**0.0375**	**0.0909**	**0.0531**	**0.0518**	**0.0710**
Improv. (%)	22.17 %	18.50%	9.29%	10.78%	9.52%	8.50%	11.64 %	13.60%	3.18%	5.10%	10.31%	5.88%	10.92 %	14.67%	2.82%	3.91%	13.10%	6.12%

Table (a)

Table (a) presents the performance of the proposed IPR in comparison to the six baselines on the three datasets. We evaluated the top-k recommendation performance with $k \in \{1, 3, 10\}$; the evaluation on a smaller k is to test the ability of first-glance recommendation and larger values of k are for "continuous scrolling" recommendation. The best results are set in boldface, and the † symbol indicates the best performing method among all the baselines. We can observe that IPR yields the best results on all evaluation metrics across all datasets with performance improvements up to 22.17% and 14.67% in recall and nDCG, respectively, which attests the effectiveness of our model. We can also observe that Compared to the performance regarding positions like k = 10, IPR performs remarkably well for positions k = 1, 3, suggesting that the proposed interaction-level modeling to include all rating magnitudes is advantageous to top-position recommendation.

Ablation Study

	Movielens-1M					
	recall@1	recall@3	recall@10	nDCG@1	nDCG@3	nDCG@10
IPR (rating>3)	0.0261	0.0645	0.1519	0.5009	0.4536	0.3869
IPR	**0.0270**	**0.0666**	**0.1564**	**0.5124**	**0.4610**	**0.3943**
Improv. (%)	3.44 %	3.25%	2.96%	2.29%	1.63%	1.91%

Table (b)

To further determine whether the proposed IPR effectively encompasses lower rating data, we twist the original setting of IPR (which considers all ratings for model training) into "IPR (rating>3)." The experiments of IPR (rating>3) follow the conventional settings in which only ratings larger than three are included into the modeling process. The two are compared in Table 4, which attests to our hypothesis that all rating data represent valuable insight and thus should be included in the modeling process