Item Similarity Mining for Multi-Market Recommendation

Jiangxia Cao
Institute of Information Engineering, CAS
School of Cyber Security, UCAS
caojiangxia@iie.ac.cn

Xin Cong
Institute of Information Engineering, CAS
School of Cyber Security, UCAS
congxin@iie.ac.cn

Tingwen Liu*
Institute of Information Engineering, CAS
School of Cyber Security, UCAS
liutingwen@iie.ac.cn

Bin Wang
Xiaomi AI Lab,
Xiaomi Inc
wangbin11@xiaomi.com

ABSTRACT

Real-world web applications such as Amazon and Netflix often provide services in multiple countries and regions (i.e., markets) around the world. Generally, different markets share similar item sets while containing different amounts of interaction data. Some markets are data-scarce and others are data-rich and leveraging those data from similar and data-rich auxiliary markets could enhance the data-scarce markets. In this paper, we explore multi-market recommendation (MMR), and propose a novel model called M^3Rec to improve all markets recommendation simultaneously. Since items play the role to bridge different markets, we argue that mining the similarities among items is the key point of MMR. Our M^3Rec preprocess two global item similarities: intra- and inter- market similarities. Specifically, we first learn the second-order intra-market similarity by adopting linear models with closed-form solutions, and then capture the high-order inter-market similarity by the random walk. Afterward, we incorporate the global item similarities for each local market. We conduct extensive experiments on five public available markets and compare with several state-of-the-art methods. Detailed experimental results demonstrate the effectiveness of our proposed method.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Computing methodologies** → **Neural networks**.

KEYWORDS

Multi-Market Recommendation; Item Similarity; Linear Model

ACM Reference Format:
Jiangxia Cao, Xin Cong, Tingwen Liu, Bin Wang. 2022. Item Similarity Mining for Multi-Market Recommendation. In *Proceedings of the 45th Int'l ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3477495.3531839

*Tingwen Liu is the corresponding author.

SIGIR '22, July 11–15, 2022, Madrid, Spain.

ACM ISBN 978-1-4503-8732-3/22/07.
https://doi.org/10.1145/3477495.3531839

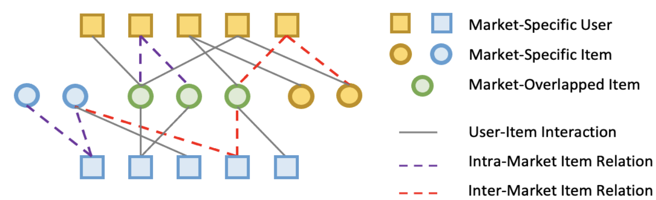

Figure 1: A simple illustration of MMR. The purple dotted lines denote second-order intra-market item relation, the red dotted lines denote high-order inter-market item relation.

1 INTRODUCTION

Many e-commerce and media applications such as Amazon [26] and Spotify [24] are deployed in multiple countries and regions, i.e., deployed in different markets. Different markets contain different amounts of interaction data between users and items where user sets are disjoint with other markets and the item sets are partially overlapped with others. Traditional academic and industrial recommender systems (RS) mainly focus on using data in a single market to train models and then serve the market, which limits the performances of data-scarce markets. Thus, it is critical to explore all market data to enhance the data-scarce markets, i.e., multi-market recommendation (MMR).

Related work. To our knowledge, the most closely related work is FOREC [3], which releases the XMarket dataset for MMR. The FOREC follows meta learning framework, it first pre-training a NeuMF [13] network on multiple markets, and then fine-tune the network on the target market.

Except it, the multi-task learning methods could also be adapted into MMR, such as Cross-Stitch [21], MMoE [19], Bi-TGCF [18] and STAR [25]. Cross-Stitch could adapt NeuMF [13] as the base encoder for each task and introduces a cross-stitch network to transfer information between base encoders. MMoE learns task relationships by sharing the mixture-of-expert base encoders across all tasks, while designing a task-specific gating network to aggregate encoder outputs to optimize each task. Bi-TGCF exploits LightGCN [12] as the base encoder to capture the collaborative filtering signal for each task and introduces a feature transfer layer to connect base encoders. STAR devises an encoder topology structure, and utilizes a task-shared base encoder to enhance all task-specific base encoders.

Our contributions. Although the above methods can be used in MMR, they ignore that overlapped items play the central role to bridge different markets. Since data-rich markets and data-scarce markets share some overlapped items in MMR, therefore data-rich markets could provide more co-occurrence information to learn overlapped items. Mining the similarity between items could make the well-learned overlapped items supported by data-rich markets enhance similar items in data-scarce markets. We argue that the key point of MMR is how to mine the multi-market item similarity. In this work, we propose a simple-yet-effective model $\mathbf{M^3Rec}$ to Mine item similarity for Multi-Market Recommendation. As shown in Figure 1, we consider two level item pair similarities: the intra-market similarity to capture the second-order item-item relation and the inter-market similarity to capture the higher-order item-item relation.

For the intra-market level similarity: since users often interact with highly coherent and consistent items, we utilize EASE^R [27] to learn the related item pairwise pattern for modeling second-order item-item relation. In this way, items interacted by the same user are encouraged to be more similar than other non-interacted items. Nevertheless, the second-order relation is insufficient since the market-specific item pairs were ignored, e.g., market-A-specific item should require at least 4 hops to connect the market-B-specific item. It motivates us to further consider higher-order inter-market item-item relations.

For the inter-market level similarity: to capture the high-order similarity of items in different markets, we exploit the idea of random walk [14] to sample various multi-market item sequences to learn item-item relation for all markets. Specifically, we first construct a global item co-occurrence graph, which could connect different markets by sampling different item sequences. Then, we leverage the skip-gram algorithm [20] with the negative sampling method to learn the robust global item embedding.

In the implementation, we first pre-train the intra-market similarity by EASE^R [27] with its closed-form solutions and the inter-market similarity by random walk. Then, we leverage them as prior to fine-tune all local market recommendation. In MMR, the market-overlapped item may have different embeddings in different markets since various biases and distributions of markets. In this work, we think an item should express nearness embeddings even in different markets. The reason is that item embeddings indicate the item intrinsic properties such as category and usage. Therefore, we further learn a L_2-norm penalized small-scale local item embedding for each item.

Our main contributions are summarized as follows:

- We introduce a fresh perspective to solve MMR by mining the item similarity, and propose a novel method M^{3}Rec.
- We consider two level item similarities to enhance MMR. The intra-market similarity is preprocessed by the linear model with closed-form solutions, while the inter-market similarity is preprocessed by the random walk model.
- We empirically evaluate M^{3}Rec on five public available markets and show its superior performance.

2 PRELIMINARY

2.1 Problem Statement

Assuming there are existing M parallel markets $\{\mathcal{M}^1, \dots, \mathcal{M}^M\}$ with similar item sets. Let $\mathcal{M}^m = (\mathcal{U}^m, \mathcal{V}^m, \mathcal{E}^m, \mathbf{X}^m)$ denote the m-th market data, where $\mathcal{U}^m, \mathcal{V}^m, \mathcal{E}^m$ are the user set, the item set, the edge set, and $\mathbf{X}^m \in \{0,1\}^{|\mathcal{U}^m| \times |\mathcal{V}^m|}$ describes the binary user-item interaction matrix. By merging all markets data, we further introduce a global market $\mathcal{M} = (\mathcal{U}, \mathcal{V}, \mathcal{E}, \mathbf{X})$ to capture the intra- and inter- market item similarity. Given those market data, our goal is to enhance predictive performance for all markets.

2.2 Linear Item Pairwise Model

The key idea of linear item pairwise model [8–10, 28, 29] is to learn second-order item-item relevance matrix $\mathbf{B} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ from observed user-item interaction matrix $\mathbf{X} \in \{0,1\}^{|\mathcal{U}| \times |\mathcal{V}|}$. Formally, the linear item-item model could make recommendations as follows:

$$\hat{\mathbf{X}} = \mathbf{X}\mathbf{B}, \tag{1}$$

where $\hat{\mathbf{X}} \in \mathcal{R}^{|\mathcal{U}| \times |\mathcal{V}|}$ is the predictive recommendation result. SLIM [22] is a pioneer work to estimate $\mathbf{B}$, it minimizes the least squares under constraints that all entries in $\mathbf{B}$ are non-negative and zero diagonal:

$$\begin{gathered}\underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{X} \cdot \mathbf{B}\|_F^2 + \lambda \|\mathbf{B}\|_1 + \lambda_F \|\mathbf{B}\|_F^2 \\ \texttt{s.t. diag}(\mathbf{B}) = 0, \quad \mathbf{B} \geq 0,\end{gathered} \tag{2}$$

where $\|\cdot\|_1$ and $\|\cdot\|_F$ denote the L_1 norm and Frobenius norm, λ and λ_F are hyper-parameters. The constraint of zero diagonal $\texttt{diag}(\mathbf{B}) = 0$ is crucial to avoid $\mathbf{B} = \mathbf{I}$ to prevent the model from overfitting problem [29]. Although SLIM provides a brief objective, but it is notorious to train. Recently, EASE^R [27] simplify SLIM by removing the L_1 norm and the non-negative constraint on $\mathbf{B}$ as:

$$\underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{X} \cdot \mathbf{B}\|_F^2 + \lambda_F \|\mathbf{B}\|_F^2 \quad \texttt{s.t. diag}(\mathbf{B}) = 0. \tag{3}$$

As discussed in [27], the above straightforward objective could be minimized in closed-form solution by Lagrange multipliers:

$$\hat{\mathbf{B}} = \mathbf{I} - \hat{\mathbf{C}} \cdot \texttt{diagMat}\big(\mathbf{1} \oslash \texttt{diag}(\hat{\mathbf{C}})\big), \text{ where } \hat{\mathbf{C}} = (\mathbf{X}^\top \mathbf{X} + \lambda_F \mathbf{I})^{-1} \tag{4}$$

where $\hat{\mathbf{B}}$ is the predictive matrix, $\texttt{diagMat}(\cdot)$ denotes a diagonal matrix, $\mathbf{1}$ is a vector of ones and $\oslash$ means element-wise division.

2.3 Random Walk Model

The random walk models [11, 23, 30] aim at capturing the transition relations among the set of nodes, a.k.a. the item set $\mathcal{V}$. Given the **items co-occurrence weighted matrix** $\mathbf{A} = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$, for the standard random walk process, the transition probability matrix $\mathbf{P}$ is defined:

$$\mathrm{p}(v_j | v_i) = \frac{\mathbf{A}_{ij}}{\texttt{deg}(v_i)} = \mathbf{P}_{ij}, \quad \mathbf{P} = \mathbf{D}^{-1}\mathbf{A}, \tag{5}$$

where $\mathbf{D} = \texttt{diagMat}([\deg(v_1), \deg(v_2), \dots]) \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ is the degree matrix, $\deg(\cdot)$ is the degree function, $\mathrm{p}(v_j|v_i)$ is the probability from node v_i jumping to node v_j. Skip-gram algorithm [20] provides a simple way to model $\mathbf{P}$, for each item sequence $(v_i)_{i=1}^T$:

$$\operatorname{argmax} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \log \hat{\mathrm{p}}(v_j | v_i), \text{where } \hat{\mathrm{p}}(v_j|v_i) = \frac{\exp(\boldsymbol{v}_i^\top \bar{\boldsymbol{v}}_j)}{\sum_{v_k \in \mathcal{V}} \exp(\boldsymbol{v}_i^\top \bar{\boldsymbol{v}}_k)}, \tag{6}$$

where $\boldsymbol{v}, \bar{\boldsymbol{v}} \in \mathbb{R}^d$ are training target and context embedding of item.

3 METHODOLOGY

Given all parallel local markets and global market interaction data, we first pre-process two global item similarities by linear item model and random walk model. Afterward, we exploit them to make all local market recommendations simultaneously.

3.1 Global Item Similarity Mining

3.1.1 Intra-Market Item Similarity. Given the global user-item interaction matrix $\mathbf{X}$, we follow the assumption of EASE^R in Eq.(3-4) to obtain the zero-diagonal second-order item-item relevance matrix $\hat{\mathbf{B}}^{\text{Glob}}$. Actually, the relevance matrix $\hat{\mathbf{B}}^{\text{Glob}}$ is not directly reflecting the probability among items, i.e., each element is non-negative and each row summation is 1. To satisfy the constraint and accelerate computation, we first filter the smaller elements as zeros by threshold ϵ, and then normalize the matrix as:

$$\mathbf{B}^{\text{Glob}} = \texttt{Normalize}(\hat{\mathbf{B}}^{\text{Glob}}_{\geq\epsilon}), \text{ where } \hat{\mathbf{B}}^{\text{Glob}}_{\geq\epsilon} = \texttt{Filter}(\hat{\mathbf{B}}^{\text{Glob}}, \epsilon) \quad (7)$$

where the $\mathbf{B}^{\text{Glob}}$ denotes intra-market item similarity matrix[1]. In our experiments, compared with $\hat{\mathbf{B}}^{\text{Glob}}$, we find our modified matrix $\mathbf{B}^{\text{Glob}}$ does not degenerate the prediction result under the $0.02 > \epsilon$.

3.1.2 Inter-Market Item Similarity. Besides the intra-market similarity, we also consider the inter-market similarity. As discussed in § 2.3, random walk models could implicitly learn the item embedding to capture the item correlation by Eq.(6). Since we already obtain the intra-market item similarity matrix $\mathbf{B}^{\text{Glob}}$, thus we can explicitly incorporate it to train item embedding:

$$\mathbf{V} = \mathbf{B}^{\text{Glob}}\mathbf{V}^{\text{Init}} + \mathbf{V}^{\text{Init}}, \quad \bar{\mathbf{V}} = \mathbf{B}^{\text{Glob}}\bar{\mathbf{V}}^{\text{Init}} + \bar{\mathbf{V}}^{\text{Init}}, \quad (8)$$

where $\mathbf{V}^{\text{Init}}, \bar{\mathbf{V}}^{\text{Init}} \in \mathbb{R}^{|\mathcal{V}|\times d}$ are initialized target and context parameter matrix of item. Besides, to avoid enormous computation cost of $\hat{\text{p}}(v_j|v_i)$ in Eq.(6), we leverage negative sampling technique to replace it for large-scale item training:

$$\hat{\text{p}}(v_j|v_i) = \texttt{Sigmoid}(\boldsymbol{v}_i^\top \bar{\boldsymbol{v}}_j) \prod_{v_k \in \mathcal{N}} \texttt{Sigmoid}(\boldsymbol{v}_i^\top \bar{\boldsymbol{v}}_k), \quad (9)$$

where $\boldsymbol{v} \in \mathbf{V}, \bar{\boldsymbol{v}} \in \bar{\mathbf{V}}, \mathcal{N}$ is the negative samples set. Afterward, we keep the same objective in Eq.(6) to train embedding $\mathbf{V}$ to reflect inter-market similarity, we denote it as $\mathbf{V}^{\text{Glob}}$ in following section.

3.2 Local Market Recommendation

After obtaining the preprocessed intra-market item similarity matrix $\mathbf{B}^{\text{Glob}}$ and inter-market item embedding $\mathbf{V}^{\text{Glob}}$, we then conduct the market adaptation operation for each local market.

For simplicity, we take the m-th market $\mathcal{M}^m$ as example. Since the local market only including a partial items $\mathcal{V}^m \in \mathcal{V}$, we first tailor $\mathbf{B}^{\text{Glob}}, \mathbf{V}^{\text{Glob}}$ as $\mathbf{B}^{\text{Glob}}_m \in \mathbb{R}^{|\mathcal{V}^m|\times|\mathcal{V}^m|}$, $\mathbf{V}^{\text{Glob}}_m \in \mathbb{R}^{|\mathcal{V}^m|\times d}$ for adaptation[2]. Intuitively, the same item may have different embeddings in different markets, but should express nearness embeddings. The reason is that item embeddings indicate the item intrinsic properties such as category and usage, thereby we have:

$$\mathbf{V}^m = \mathbf{V}^{\text{Glob}}_m + \mathbf{V}^{\text{Loca}}_m, \quad (10)$$

[1]As discussed before, the second-order similarity neglects market-specific item pairs.
[2]Note that we also normalize the tailored matrix $\mathbf{B}^{\text{Glob}}_m$.

Table 1: Statistics of real-world anonymized markets, #Rate denotes the percent of its items shared by other markets.

	M1	M2	M3	M4	M5
$\lvert\mathcal{U}\rvert$	7,109	2,697	3,328	5,482	6,466
$\lvert\mathcal{V}\rvert$	2,198	1,357	1,245	2,917	9,762
#Train	48,302	19,615	23,367	41,226	77,173
#Valid	3,534	1,375	1,639	2,706	3,207
#Test	3,575	1,322	1,689	2,776	3,259
#Rate	75.97%	84.89%	89.23%	72.78%	27.42%

where $\mathbf{V}^{\text{Loca}}_m$ is initialized parameter matrix of $\mathcal{M}^m$. Besides, $\mathbf{V}^{\text{Loca}}_m$ is forced to be small-scale, by penalizing the L_2-norm of parameters. For the user embeddings, inspired by the neighboring aggregation idea [4, 5, 12], we learn them from $\mathbf{V}^m$ as follows:

$$\mathbf{U}^m = \hat{\mathbf{X}}^m\mathbf{V}^m + \mathbf{U}^{\text{Loca}}_m, \text{ where } \hat{\mathbf{X}}^m = \texttt{Normalize}(\mathbf{X}^m\mathbf{B}^{\text{Glob}}_m), \quad (11)$$

where $\hat{\mathbf{X}}^m \in \mathbb{R}^{|\mathcal{U}^m|\times|\mathcal{V}^m|}$ is adaptive user-item matrix incorporating intra-market similarity. Finally, we measure the `Cosine` distance $\hat{y}^m_{ij}$ between user $u_i \in \mathcal{U}^m$ and item $v_j \in \mathcal{V}^m$ for training:

$$\begin{aligned} \hat{y}^m_{ij} &= \texttt{Cosine}(\boldsymbol{u}^m_i, \boldsymbol{v}^m_j), \\ \mathcal{L}^m(u_i, v_j) &= (1 - \hat{y}^m_{ij}) + \frac{1}{|\mathcal{N}|}\sum_{v_k \in \mathcal{N}} \texttt{max}(0, \hat{y}^m_{ik} - \gamma) \end{aligned} \quad (12)$$

where $\boldsymbol{u}^m_i \in \mathbf{U}^m, \boldsymbol{v}^m_j \in \mathbf{V}^m, \mathcal{N}$ denotes the negative samples set, γ is the margin to filter negative samples.

3.3 Model Optimization

Our model could be optimized over the M markets simultaneously:

$$\mathcal{L} = \sum_{m=1}^{M} (\mathcal{L}^m + \lambda\|\mathbf{V}^{\text{Loca}}_m\|_2), \quad (13)$$

where the $\|\cdot\|_2$ is the L_2 norm, λ is the factor to control the local embedding to be small-scale. We train the model by minimizing $\mathcal{L}$ through Adam stochastic gradient descent [16] over the shuffled mini-batches.

4 EXPERIMENTS

4.1 Experimental Setup

4.1.1 Datasets. Following FOREC [3], we evaluate our model on the subset electronics domain data from XMarket dataset[3], which includes several real-world markets, e.g., United Kingdom, Germany, France, Canada, United States and so on. To be specific, we exploit the public five pre-processed anonymized markets [2, 7, 32, 33] of electronics XMarket data[4] to train our model. For evaluation, we randomly select 50% users for validation and the other 50% users for test. The concrete statistics of users, items, interactions and item overlapping rates are listed in Table 1.

[3]https://xmrec.github.io/
[4]https://github.com/hamedrab/wsdm22_cup_xmrec/tree/main/DATA

Table 2: Performance comparison of different methods.

	Models	MRR					NDCG@10					HR@10				
		M1	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**	**M1**	**M2**	**M3**	**M4**	**M5**
Mix	NeuMF	41.40	28.50	29.17	27.19	8.87	46.31	34.84	36.57	29.83	9.01	62.73	55.60	60.40	40.33	12.26
	LightGCN	43.85	26.37	27.14	28.86	12.58	48.30	32.70	34.22	31.03	13.16	64.73	52.13	56.26	41.14	17.13
	Random Walk	42.87	25.51	27.83	27.51	13.32	48.05	31.10	34.89	30.08	14.07	64.66	50.20	57.53	39.20	17.73
	EASE^R	48.96	30.05	31.70	33.45	16.58	54.95	37.92	40.13	36.63	16.93	70.80	57.40	63.60	45.13	19.13
Multi	Cross-Stitch	44.37	32.61	34.13	27.47	11.75	49.15	36.65	39.13	29.16	12.51	64.46	54.06	59.46	38.93	17.06
	MMoE	45.36	34.78	36.48	28.55	11.29	48.98	38.71	41.30	30.16	11.79	65.73	56.26	61.53	38.60	16.66
	Bi-TGCF	45.29	27.31	28.55	30.18	13.78	50.46	33.43	35.77	32.76	14.42	66.86	53.46	58.73	42.26	17.86
	STAR	41.66	30.87	33.43	27.30	11.10	46.57	35.25	37.09	29.84	12.48	62.93	54.89	60.80	40.20	16.60
	FOREC	48.26	35.89	36.57	31.61	12.50	52.05	40.03	41.88	33.52	13.19	65.06	58.42	64.13	41.60	17.46
	M^3Rec	**50.24**	**36.41**	**38.75***	**35.00**	**17.93***	**55.83**	**40.04**	**43.35***	**37.98***	**18.63***	**73.13***	**60.86***	**66.53***	**48.46***	**22.66***

* indicates that the improvements are statistically significant for $p < 0.05$ judged with the runner-up result in each case by paired t-test.

Table 3: Performance of different model variants.

Model variants	MRR				
	M1	**M2**	**M3**	**M4**	**M5**
Complete	50.24	36.41	38.75	35.00	17.93
w/o intra-market similarity	47.48	32.40	33.01	33.23	15.31
w/o inter-market similarity	48.37	33.16	35.66	33.79	16.73
w/o both similarities	46.47	29.10	31.21	31.92	13.10

Table 4: Performance of two item groups.

Item groups	MRR				
	M1	**M2**	**M3**	**M4**	**M5**
Market-overlapped item	46.15	34.72	36.63	32.52	26.03
Market-specific item	61.32	37.42	45.31	42.77	11.50
Market-overlapped item*	41.47	28.62	27.90	28.37	14.63
Market-specific item*	60.00	30.86	44.36	38.13	9.51

4.1.2 Baselines. We compare M^3Rec with several baselines which could be categorized into four classes: (1) Traditional methods: NeuMF [13] and LightGCN [12], (2) Item-based methods: Random Walk [1] and EASE^R [27], (3) Multi-task methods: Cross-Stitch [21], MMoE [19], Bi-TGCF [18] and STAR [25], (4) Multi-market method: FOREC [3]. The main details of these methods have already been introduced in related work and preliminary.

4.1.3 Evaluation Protocol. Following FOREC [3], we also leverage the leave-one-out method [6] to calculate the recommendation performance. To guarantee unbiased evaluation, we follow Rendle's literature [17] to calculate 1000 result list for each test case (containing 999 negative items and 1 positive item). Afterward, we adopt three widely-used metrics MRR (Mean Reciprocal Rank [31]), NDCG (Normalized Discounted Cumulative Gain [15]) and HR (Hit Ratio) to show performance on the top-10 ranking results.

4.1.4 Parameter Settings. For a fair comparison, we keep the same hyper-parameters setting for each method: the embedding size d is fixed as 128, the dropout rate is fixed as 0.3, the learning rate is set as 0.001, the batch size is fixed as 1024, the training epoch is set as 100 to get the best result, the negative sampling number $|\mathcal{N}|$ is fixed as 10, and Adam [16] optimizer is used to update all parameters. Besides, the hyper-parameter λ_F of EASE^R is selected from 10 to 50 with step length 5, the window size of random walk is fixed as 10, the hidden neural network structure and the latent factor dimension of base encoder is set as $[256 \rightarrow 512 \rightarrow 1024 \rightarrow 128]$, the deep of base GNN encoder is selected from {1,2,3}, the L_2 regularization coefficient λ of our method is chosen from $\{1e^{-6},1e^{-7},1e^{-8}\}$, the filtering threshold ϵ is selected from 0.01 to 0.02 with step length 0.001, and the margin γ of our method is selected from{0.1, 0.2, 0.3, 0.4}. In the following section, we report M^3Rec results under $\lambda = 1e^{-7}$, $\epsilon = 0.01$ and $\gamma = 0.2$ by default.

4.2 Performance Comparisons

Table 2 shows the recommendation performance of M^3Rec against other baselines in terms of MRR, NDCG@10 and HR@10 on the five markets. To be specific, the 'Mix' means that we directly merge interaction data of markets and train a model to predict all markets. The 'Multi' denotes that the model leverages different modules for different markets and transfer information across different market modules. From Table 2, we have several insightful observations:

For 'Mix' models: (1) The item-based method EASE^R largely outperforms other models. This observation validates that the item similarity is crucial to make predictions. (2) LightGCN and Random Walk show consistently improve than NeuMF in markets **M1**, **M4** and **M5**, but gain worsen results in markets **M2** and **M3**. The reason might be that the data-scarce market **M2** and **M3** have only a small amount of interaction data than others (see in Table 1). Therefore, the imbalance data issue with jointly training strategy would cause the prediction result to be dominated by the other data-rich markets.

For 'Multi' models: (1) Comparing with NeuMF and LightGCN, their extension methods are shown satisfactory improvements in all markets, which indicates that transferring information in different market modules is a promising way to enhance MMR. (2) Our

M^{3}Rec achieves the best metrics in all markets, which shows that mining item similarity is the key point to enhance MMR.

4.3 Discussion of Model Variants

To investigate the effectiveness of our model components, we conduct experiments of three M^{3}Rec variants in Table 3. Note that the 'w/o' denotes removing the corresponding item similarity from the complete model. From it, we can observe that: (1) The variants without any item similarities significantly decline on MRR metric, which demonstrates that the intra- and inter- market level similarities are both beneficial for our model. (2) The impact of intra-similarities tend to be more significant than inter- level. We suppose the reason is that the intra- similarity could generate robust user-item correlations to guide better user embeddings.

4.4 Effectiveness of Item Similarity

In this section, to analyze the effect of item similarity for MMR, we further conduct experiments on two different item groups from the validation set: the market-overlapped and market-specific items. We report the performance in Table 4, and the '*' denotes that the result is obtained from the model variant without both item similarities. From the table, we have the following observations: (1) After leveraging the item similarities, the recommendation performance of both item groups shows largely improvement, which demonstrates that mining item similarities not only enhances the market-overlapped items, but also benefits the market-specific items. Meanwhile, compared with market-specific items, the improvement of market-overlapped items is more significant. This is because the overlapped items have more interactions across markets to learn robust embeddings. (2) It is surprising that the market-specific items have better recommendation results than market-overlapped items in the front four markets. After our statistics, we think the reason might be the data distribution. Except the **M5** market, we find the average interaction number of market-specific items is more than market-overlapped items, which means that the market-specific items are more popular in these markets. Thus, it is easy for models to learn to recommend these popular items. However, as shown in Table 1, the overlapping rate reflects that the number of market-specific items is much less than the market-overlapped items in the first four markets. In this case, the higher overlapping rate indicates that it is urgent to enhance the overlapped item performance by transferring information from other data-rich markets.

5 CONCLUSIONS

In this paper, we investigate the multi-market recommendation (MMR) and propose a novel approach M^{3}Rec. Specifically, we first provide a fresh perspective to solve MMR by mining the item similarity, which plays the central role to bridge different markets. Then, we consider two global item similarities for MMR, the intra-market item similarity is learned by linear model EASE^R and the inter-market item similarity is generated by random walk. Finally, we fuse the two global item similarities in our local market recommendation. Extensive experiments demonstrate that M^{3}Rec outperforms current state-of-the-art methods, and detailed analyses demonstrate the effectiveness of our model components.

ACKNOWLEDGEMENT

We would like to thank Yanzeng Li, Jiawei Sheng, Shiyao Cui and Shicheng Wang for their helpful discussions and feedback of this work. We would also like to thank anonymous reviewers for their constructive comments. This work was supported by the National Key Research and Development Program of China under Grant No.2021YFB3100600, the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No.XDC02040400, and the Youth Innovation Promotion Association of CAS under Grant No.2021153.

REFERENCES

[1] Oren Barkan and Noam Koenigstein. 2016. Item2vec: neural item embedding for collaborative filtering. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*.
[2] Cesare Bernardis. 2022. Multi-stage Ensemble Model for Cross-market Recommendation. *ArXiv* (2022).
[3] Hamed Bonab, Mohammad Aliannejadi, Ali Vardasbi, Evangelos Kanoulas, and James Allan. 2021. Cross-Market Product Recommendation. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[4] Jiangxia Cao, Xixun Lin, Xin Cong, Shu Guo, Hengzhu Tang, Tingwen Liu, and Bin Wang. 2021. Deep Structural Point Process for Learning Temporal Interaction Networks. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*.
[5] Jiangxia Cao, Xixun Lin, Shu Guo, Luchen Liu, Tingwen Liu, and Bin Wang. 2021. Bipartite Graph Embedding via Mutual Information Maximization. In *ACM International Conference on Web Search and Data Mining (WSDM)*.
[6] Jiangxia Cao, Jiawei Sheng, Xin Cong, Tingwen Liu, and Bin Wang. 2022. Cross-Domain Recommendation to Cold-Start Users via Variational Information Bottleneck. In *IEEE International Conference on Data Engineering (ICDE)*.
[7] Zeyuan Chen, He Wang, Xiangyu Zhu, Haiyan Wu, Congcong Gu, Shumeng Liu, Jinchao Huang, and Wei Zhang. 2022. A Practical Two-stage Ranking Framework for Cross-market Recommendation. In *WSDM Cup Workshop*.
[8] Minjin Choi, Jinhong Kim, Joonseok Lee, Hyunjung Shim, and Jongwuk Lee. 2021. Session-aware linear item-item models for session-based recommendation. In *The World Wide Web Conference (WWW)*.
[9] Minjin Choi, Jinhong Kim, Joonsek Lee, Hyunjung Shim, and Jongwuk Lee. 2022. S-Walk: Accurate and Scalable Session-based Recommendationwith Random Walks. In *ACM International Conference on Web Search and Data Mining (WSDM)*.
[10] Evangelia Christakopoulou and George Karypis. 2014. Hoslim: Higher-order sparse linear method for top-n recommender systems. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*.
[11] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[12] Xiangnan He, Kuan Deng, Xiang Wang, Yaliang Li, Yongdong Zhang, and Meng Wang. 2020. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In *ACM International Conference on Research on Development in Information Retrieval (SIGIR)*.
[13] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *The World Wide Web Conference (WWW)*.
[14] Zexi Huang, Arlei Silva, and Ambuj Singh. 2021. A Broader Picture of Random-walk Based Graph Embedding. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[15] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* (2002).
[16] P. Diederik Kingma and Lei Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*.
[17] Walid Krichene and Steffen Rendle. 2020. On sampled metrics for item recommendation. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[18] Meng Liu, Jianjun Li, Guohui Li, and Peng Pan. 2020. Cross Domain Recommendation via Bi-directional Transfer Graph Collaborative Filtering Networks. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[19] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[20] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Annual Conference on Neural Information Processing Systems (NeurIPS)*.
[21] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. 2016. Cross-stitch networks for multi-task learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[22] Xia Ning and George Karypis. 2011. Slim: Sparse linear methods for top-n recommender systems. In *International Conference on Data Mining (ICDM)*.
[23] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *ACM Knowledge Discovery and Data Mining (KDD)*.
[24] Kevin Roitero, Ben Carterette, Rishabh Mehrotra, and Mounia Lalmas. 2020. Leveraging behavioral heterogeneity across markets for cross-market training of recommender systems. In *Companion Proceedings of the Web Conference*.
[25] Xiang-Rong Sheng, Liqin Zhao, Guorui Zhou, Xinyao Ding, Binding Dai, Qiang Luo, Siran Yang, Jingshan Lv, Chi Zhang, Hongbo Deng, et al. 2021. One Model to Serve All: Star Topology Adaptive Recommender for Multi-Domain CTR Prediction. In *ACM International Conference on Information and Knowledge Management (CIKM)*.
[26] Brent Smith and Greg Linden. 2017. Two decades of recommender systems at Amazon. com. *IEEE International Computing* (2017).
[27] Harald Steck. 2019. Embarrassingly shallow autoencoders for sparse data. In *The World Wide Web Conference (WWW)*.
[28] Harald Steck, Maria Dimakopoulou, Nickolai Riabov, and Tony Jebara. 2020. Admm Slim: Sparse Recommendations for Many Users. In *ACM International Conference on Web Search and Data Mining (WSDM)*.
[29] Harald Steck and Dawen Liang. 2021. Negative Interactions for Improved Collaborative Filtering: Don't go Deeper, go Higher. In *ACM Conference on Recommender Systems (RecSys)*.
[30] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *The World Wide Web Conference (WWW)*.
[31] Ellen M. Voorhees. 1999. The TREC-8 Question Answering Track Report. In *Text REtrieval Conference (TREC)*.
[32] Peng Zhang, Rongxiu Gao, Changyv Li, Baoyin Liu, and Sheng Zhou. 2022. A Learning to Rank method for Cross-market Recommendation. In *WSDM Cup Workshop*.
[33] Qi Zhang, Zijian Yang, Yilun Huang, Jiarong He, and Lixiang Wang. 2022. An Effective Way for Cross-Market Recommendation with Hybrid Pre-Ranking and Ranking Models. In *WSDM Cup Workshop*.