Towards Open-World Recommendation: An Inductive Model-based Collaborative Filtering Approach

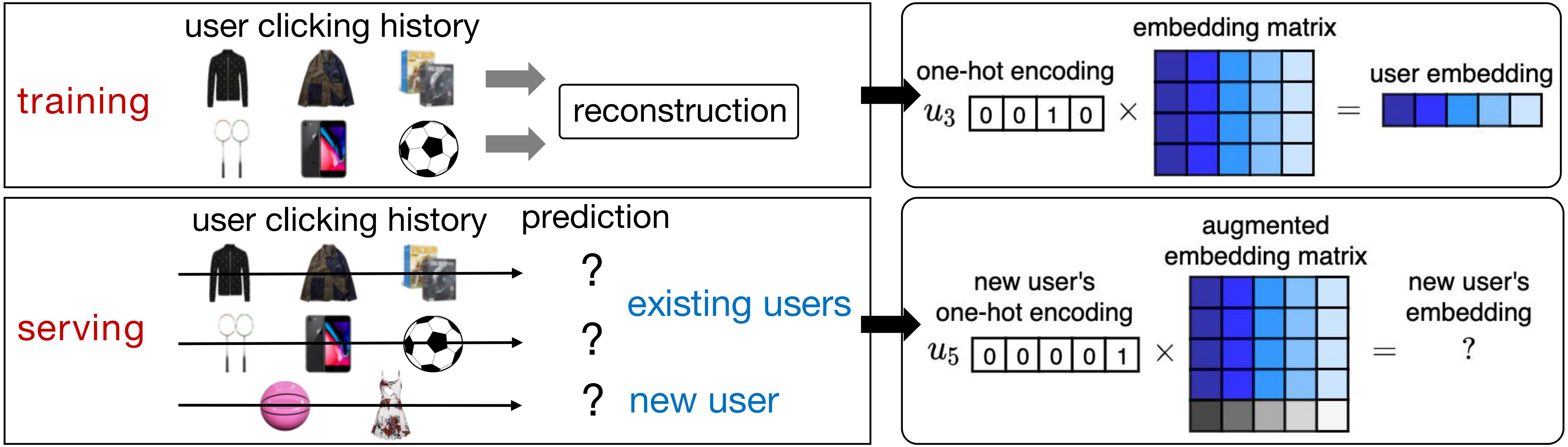
Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Junchi Yan, Hongyuan Zha
Shanghai Jiao Tong University
The Chinese University of Hong Kong, Shenzhen

Background for Recommendation

- **Model-based Collaborative Filtering ≈ Matric Factorization Model**
- **Basic idea:**

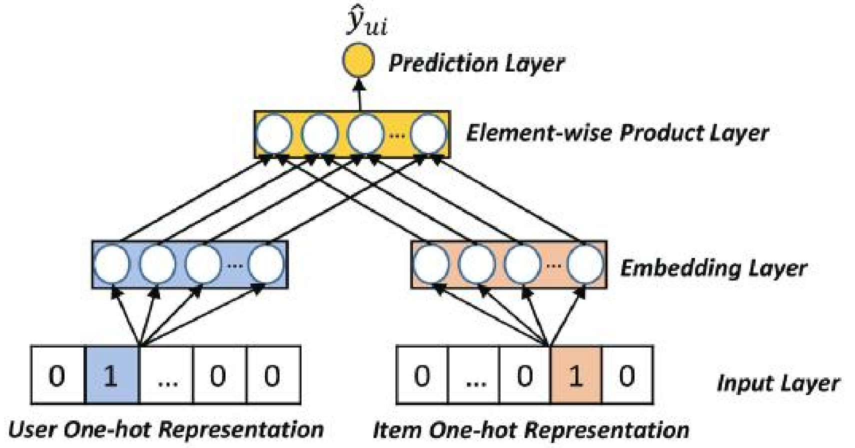
user-item rating matrix $\xrightarrow{\text{factorize}}$ { user embeddings, item embeddings } → reconstruction matrix

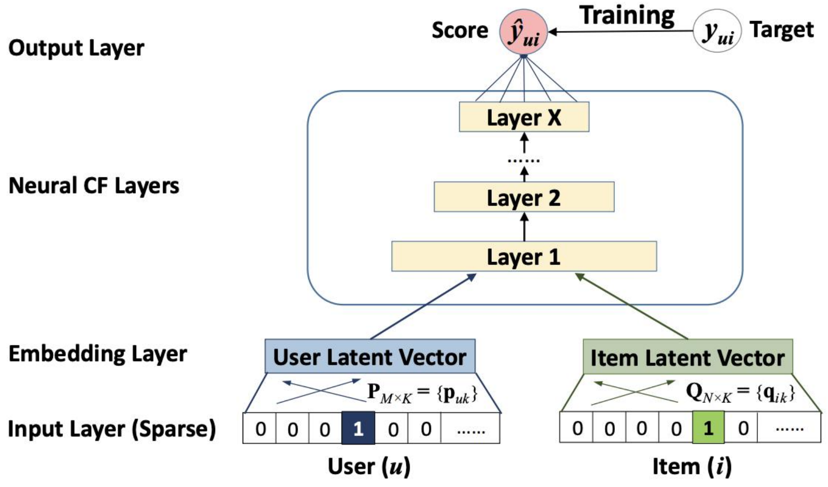
- CF models cannot handle new unseen users in open-world recommendation



Collaborative Filtering

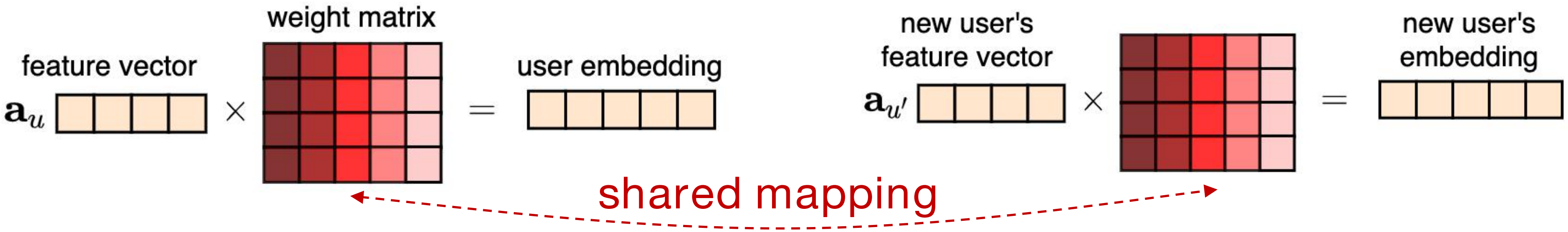
- **Formulation of CF model for RecSys:**
  - a user-item interaction matrix $R = \{r_{ui}\}_{M\times N}$
  - assume user latent factors $\mathbf{P} = \{\mathbf{p}_u\}_{M\times d}$
  - assume item latent factors $\mathbf{Q} = \{\mathbf{q}_i\}_{N\times d}$
  - consider an interaction model $\hat{r}_{ui} = f_\theta(\mathbf{p}_u, \mathbf{q}_i)$
  - target objective $\mathcal{L}(\hat{R}, R) = \sum_{(u,i)} L(\hat{r}_{ui}, r_{ui})$
- **Limitations: transductive learning**
  - cannot handle new unseen users
    - model retraining requires additional cost
    - retraining may also lead to over-fitting





Challenges for Inductive Learning

- **Inductive learning can be achieved via shared mapping**



- **Expressiveness would be sacrificed with inductive learning**

$$u_1 \xrightarrow{f_1} \mathbf{p}_{u_1} \qquad u_2 \xrightarrow{f_2} \mathbf{p}_{u_2}$$

transductive learning

pros: sufficient expressiveness

cons: fail for new users

v. s.

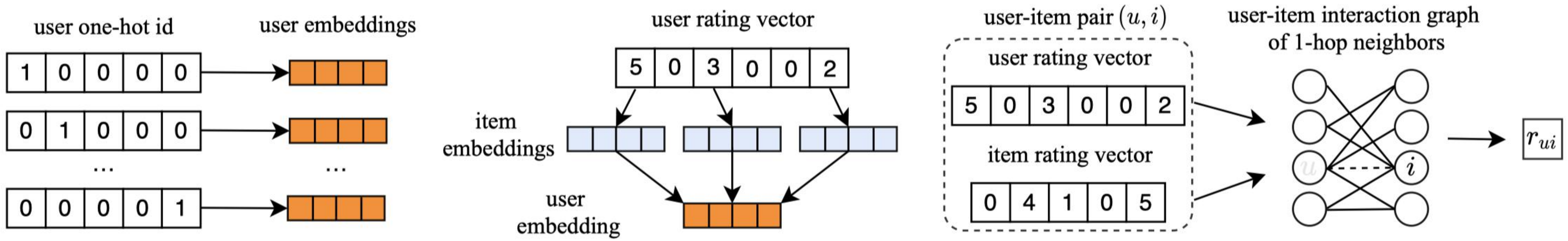
$$\mathbf{a}_{u_1} \xrightarrow{f} \mathbf{p}_{u_1} \qquad \mathbf{a}_{u_2} \xrightarrow{f} \mathbf{p}_{u_2}$$

inductive learning

pros: flexible for new users

cons: limited capacity/expressiveness

Related Works and Comparison



(a) General CF model | (b) Item-based model | (c) Local-graph-based inductive model

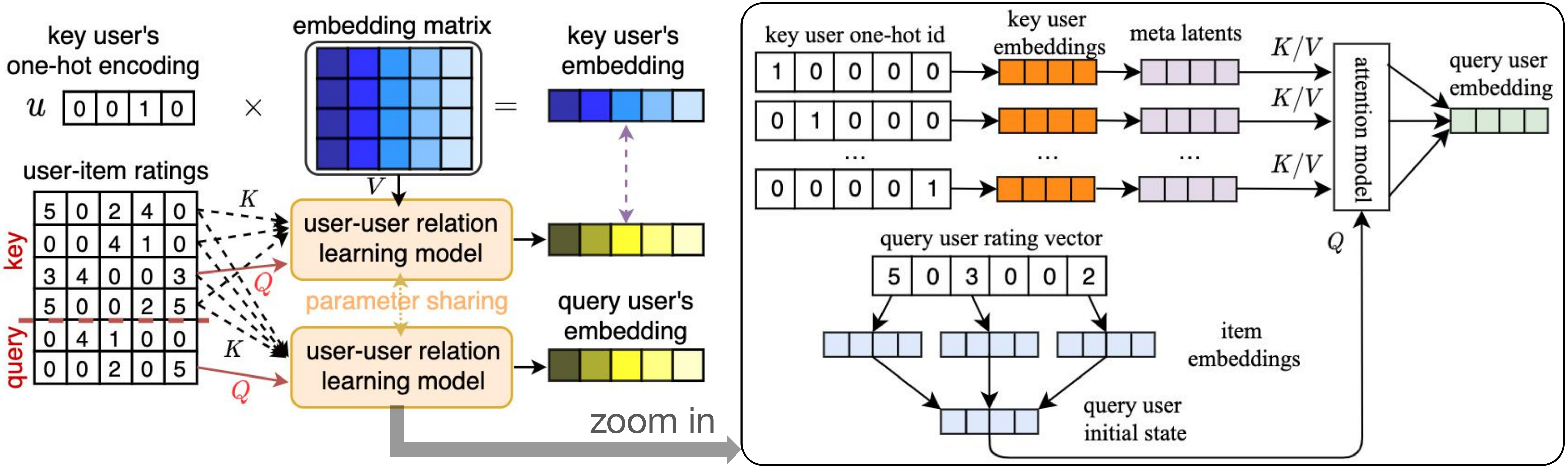
incremental learning	one-side learning	index-agnostic learning [ICLR'20]
pros:	pros:	pros:
1. advanced capacity	1. fewer parameters	1. enable inductive
2. fast training/inference	2. enable inductive	2. not require features
cons:	cons:	cons:
1. bad generalization	1. limited capacity	1. limited expressiveness
2. over-parametrization	2. user-item imbalance	2. fail for implicit feedback

Our Solutions: Inductive CF Model

❑ Basic idea:

- leverage one group of users to express another
- learn a latent graph over users
- message passing from existing users to new ones

Key insight: user preferences share underlying proximity that induces latent graphs



Our Solutions: Inductive CF Model (Cont.)

- **Partition users into two groups:** $|\mathcal{U}_k| = M_k \quad |\mathcal{U}_q| = M_q$
  - **Key users: transductive learning (traditional model)**

model: $\mathbf{P}_k = \{\mathbf{p}_u\}_{M_k \times d} \qquad \mathbf{Q} = \{\mathbf{q}_i\}_{N \times d} \qquad \hat{r}_{ui} = f_\theta(\mathbf{p}_u, \mathbf{q}_i)$

learning: $\min_{\mathbf{P}_k, \mathbf{Q}, \theta} \mathcal{D}_{\mathcal{S}_k}(\hat{R}_k, R_k)$ where $R_k = \{r_{ui}\}_{M_k \times N}$

  - **Query users: inductive learning (new model)**

model: $\tilde{\mathbf{p}}_{\mathbf{u}'} = \mathbf{c}_{u'}^\top \mathbf{P}_k \qquad c_{u'u} = \dfrac{\mathbf{e}^\top[\mathbf{W}_q \mathbf{d}_{u'} \oplus \mathbf{W}_k \mathbf{p}_u]}{\sum_{u_o \in \mathcal{U}_k} \mathbf{e}^\top[\mathbf{W}_q \mathbf{d}_{u'} \oplus \mathbf{W}_k \mathbf{p}_{u_0}]}$ where $\mathbf{d}_{u'} = \sum_{i \in \mathcal{I}_{u'}} \mathbf{q}_i$

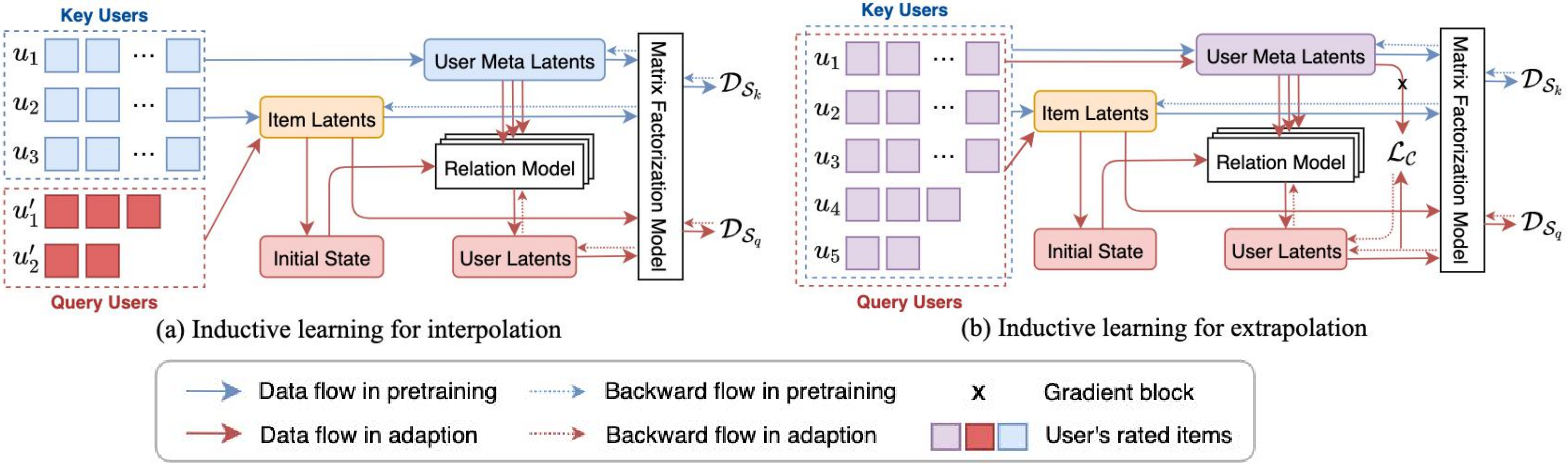
learning: $\min_{w,\theta} \mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q)$ where $R_q = \{r_{ui}\}_{M_q \times N} \qquad \hat{r}_{ui} = f_\theta(\tilde{\mathbf{p}}_{\mathbf{u}}, \mathbf{q}_i)$

objective: $\min_{w,\theta} \mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q) + \lambda \mathcal{L}_C(\mathbf{P}_k, \tilde{\mathbf{P}}_{\mathbf{k}}) \qquad \mathcal{L}_C(\mathbf{P}_k, \tilde{\mathbf{P}}_{\mathbf{k}}) = \dfrac{1}{M_q} \sum_{u \in \mathcal{U}_k} \log \dfrac{\exp(\mathbf{p}_u^\top \tilde{\mathbf{p}}_{\mathbf{u}})}{\sum_{u' \in \mathcal{U}_q} \exp(\mathbf{p}_u^\top \tilde{\mathbf{p}}_{\mathbf{u}'})}$

regularization: consistency between two estimated embeddings for one user

Our Solutions: Inductive CF Model (ont.)

- Learning procedures: pretraining + adaption
- Consider two scenarios in open-world recommendation:
  - few-shot users: pretrained model fine-tunes on new users' data
  - zero-shot users: pretrained model directly operate on new users



(a) Inductive learning for interpolation

(b) Inductive learning for extrapolation

Theoretical Analysis

- The model possesses the same representation capacity compared to matrix factorization
  - The only mild condition is that key users' latent factors span the latent space
- The generalization ability on new users depends on number of key users and training instances of new users

Theorem 1. *Assume Eq. (3) can achieve $\mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q) < \epsilon$ and the optimal $\mathbf{P}_k$ given by Eq. (1) satisfies column-full-rank, then there exists at least one solution for $\mathbf{C}$ in Eq. (2) such that $\mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q) < \epsilon$.*

$$\min_{\mathbf{P}_k, \mathbf{Q}, \theta} \mathcal{D}_{\mathcal{S}_k}(\hat{R}_k, R_k), \tag{1}$$

$$\min_{\mathbf{C}, \mathbf{Q}} \mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q), \tag{2}$$

$$\min_{\tilde{\mathbf{P}}_q, \mathbf{Q}} \mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q), \tag{3}$$

Theorem 2. *Assume 1) $\mathcal{D}$ is L-Lipschitz, 2) for $\forall \hat{r}_{u'i} \in \hat{R}_q$ we have $|\hat{r}_{u'i}| \leq B$, and 3) the L1-norm of $\mathbf{c}_{u'}$ is bounded by H. Then with probability at least $1 - \delta$ over the random choice of $\mathcal{S}_q \in ([M_q] \times [N])^{T_q}$, it holds that for any $\hat{R}_q$, the gap between $\mathcal{D}(\hat{R}_q, R_q)$ and $\mathcal{D}_{\mathcal{S}_q}(\hat{R}_q, R_q)$ will be bounded by*

$$O\left(2LHB\sqrt{\frac{2M_q \ln M_k}{T_q}} + \sqrt{\frac{ln(1/\delta)}{T_q}}\right). \tag{8}$$

Experiments Setup

❑ Dataset information:

	Dataset	# Users	#Items	# Ratings	Density	# Key/Query Users	# Training/Test Instances
explicit	Douban	3,000	3,000	0.13M	0.0152	2,131/869	80,000/20,000
	Movielens-100K	943	1,682	0.10M	0.0630	123,202/13,689	
	Movielens-1M	6,040	3,706	1.0M	0.0447	5,114/926	900,199/100,021
implicit	Amazon-Books	52,643	91,599	2.1M	0.0012	49,058/3,585	2,405,036/526,430
	Amazon-Beauty	2,944	57,289	0.08M	0.0004	780/2,164	53,464/29,440

❑ Evaluation Protocol:

- Explicit dataset: random split, RMSE & NDCG metric
- Implicit dataset: leave-last-out, AUC & NDCG metric, negative sampling

❑ Comparison: CF models, inductive models, feature-based models

Experiments Setup

- **Implementation:**
  - IDCF-NN: feedforward neural network as predictor

$$f_\theta(\mathbf{p}_u, \mathbf{q}_i) = \frac{(\mathbf{p}_u^\top \mathbf{q}_i + nn([\mathbf{p}_u \| \mathbf{q}_i \| \mathbf{p}_u \odot \mathbf{q}_i]))}{2} + b_u + b_i$$

  - IDCF-GC: graph convolution network as predictor

$$\mathbf{m}_{u,m} = ReLU\Big(\frac{1}{|\mathcal{N}_{u,m}|} \sum_{i \in \mathcal{N}_{u,m}} \mathbf{W}_{q,m} \mathbf{q}_i\Big)$$

$$\mathbf{n}_{i,m} = ReLU\Big(\frac{1}{|\mathcal{N}_{i,m}|} \sum_{u \in \mathcal{N}_{i,m}} \mathbf{W}_{p,m} \mathbf{p}_u\Big)$$

$$f(\mathbf{p}_u, \mathbf{q}_i, \{\mathbf{p}_u\}_{u \in \mathcal{N}_i}, \{\mathbf{q}_i\}_{i \in \mathcal{N}_u}) = nn'([\mathbf{p}_u \odot \mathbf{q}_i \| \mathbf{p}_u \odot \mathbf{m}_u \| \mathbf{n}_i \odot \mathbf{q}_i \| \mathbf{n}_i \odot \mathbf{m}_u]) + b_u + b_i$$

Experiments

❑ Comparison results for explicit feedback:

- For few-shot query users, very competitive results as inductive models and very close test performance to transductive models
- For zero-shot new users, significantly outperform SOTA inductive models

Method	Inductive	Feature	Douban				ML-100K				ML-1M			
			RMSE		NDCG		RMSE		NDCG		RMSE		NDCG	
			All	FS	All	FS	All	FS	All	FS	All	FS	All	FS
PMF	No	No	0.737	0.718	0.939	0.954	0.932	1.003	0.858	0.843	0.851	0.946	0.919	**0.940**
NNMF	No	No	0.729	**0.705**	0.939	0.952	0.925	0.987	0.895	0.878	0.848	**0.940**	0.920	0.937
GCMC	No	No	0.731	0.706	0.938	**0.956**	0.911	0.989	0.900	**0.886**	**0.837**	0.947	0.923	0.939
NIMC	Yes	Yes	0.732	0.745	0.928	0.931	1.015	1.065	0.832	0.824	0.873	0.995	0.889	0.904
BOMIC	Yes	Yes	0.735	0.747	0.923	0.925	0.931	1.001	0.828	0.815	0.847	0.953	0.905	0.924
F-EAE	Yes	No	0.738	-	-	-	0.920	-	-	-	0.860	-	-	-
IGMC	Yes	No	**0.721**	0.728	-	-	**0.905**	0.997	-	-	0.857	0.956	-	-
IDCF-NN (ours)	Yes	No	0.738	0.712	0.939	**0.956**	0.931	0.996	0.896	0.880	0.844	0.952	0.922	**0.940**
IDCF-GC (ours)	Yes	No	0.733	0.712	**0.940**	**0.956**	**0.905**	**0.981**	**0.901**	0.884	0.839	0.944	**0.924**	**0.940**

Method	Douban		ML-100K		ML-1M	
	RMSE	NDCG	RMSE	NDCG	RMSE	NDCG
NIMC	0.766	0.921	1.089	0.864	1.059	0.883
BOMIC	0.764	0.920	1.088	0.859	1.057	0.879
FISM	1.910	0.824	1.891	0.760	2.283	0.771
MultVAE	2.783	0.823	2.865	0.758	2.981	0.792
IGMC	0.743	-	1.051	-	0.997	-
IDCF-NN	0.749	**0.955**	1.078	0.877	0.994	0.941
IDCF-GC	**0.723**	**0.955**	**1.011**	**0.881**	**0.957**	**0.942**

Lower RMSE and higher NDCG are better

Experiments

❑ Comparison results for implicit feedback:

- For few-shot query users, achieve SOTA results
- For zero-shot new users, significantly improvement

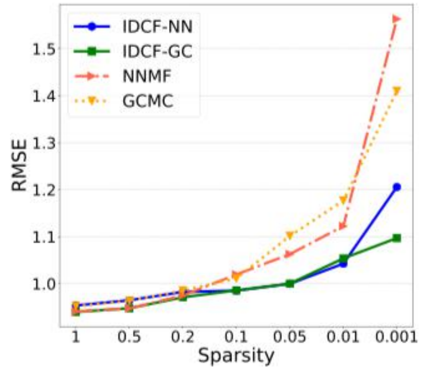
Method	Amazon-Books				Amazon-Beauty			
	AUC		NDCG		AUC		NDCG	
	Query	New	Query	New	Query	New	Query	New
PMF	0.917	-	0.888	-	0.779	-	0.769	-
NNMF	0.919	-	0.891	-	0.790	-	0.763	-
NGCF	0.916	-	0.896	-	0.793	-	0.775	-
PinSAGE	0.923	-	0.901	-	0.790	-	0.775	-
FISM	-	0.752	-	0.792	-	0.613	-	0.678
MultVAE	-	0.738	-	0.701	-	0.644	-	0.679
IDCF-NN	**0.944**	0.939	**0.928**	0.920	0.792	0.750	**0.783**	0.774
IDCF-GC	0.938	**0.946**	0.921	**0.930**	**0.801**	**0.791**	0.772	**0.791**

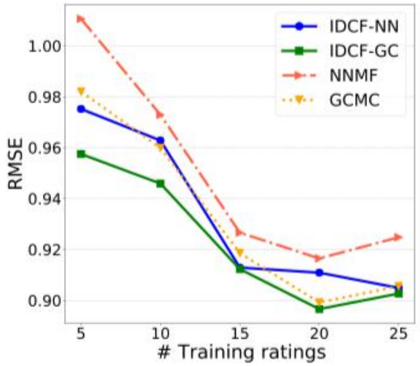
Higher AUC and higher NDCG are better

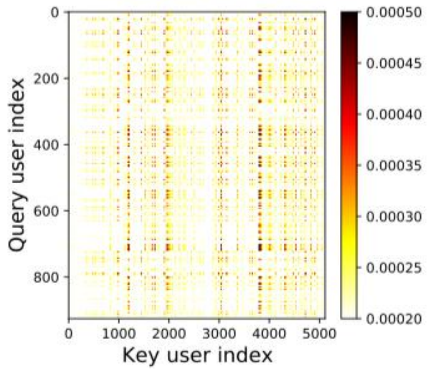
Experiments (cont.)

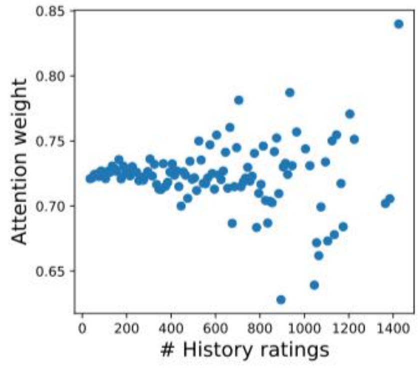
❑ Further discussions:

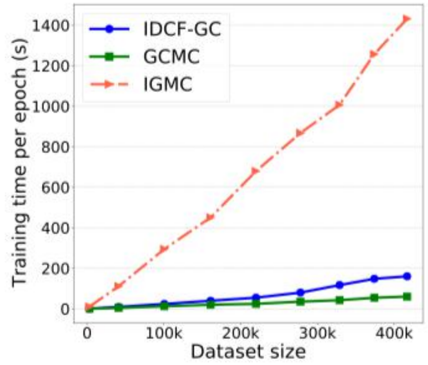
- Our model can exceed transductive models w,r,t, RMSE when users' training/historical ratings are sparse
- There exist informative key users that contribute to most of capacity. Key users with more historical ratings tend to be more important
- The training time scales linearly w.r.t. dataset size











Conclusions

❑ **Our contributions:**

- propose a new inductive representation model for CF problem
- guarantee equivalent capacity to MF and can handle new users on-the-fly
- competitive results on few-shot and SOTA performance on zero-shot users

Thanks for listening!

Paper: https://arxiv.org/abs/2007.04833
Code: https://github.com/qitianwu/IDCF