ACMmultimedia
Nice, France 21-25 October 2019

Collaborative Preference Embedding against Sparse Labels

Shilong Bao[1,2], Qianqian Xu[3], Ke Ma[1,2], Zhiyong Yang[1,2], Xiaochun Cao[1,2], Qingming Huang[1,3]

MM 2019

[1]University of Chinese Academy of Sciences, China
[2]Institute of Information Engineering, Chinese Academy of Sciences
[3]Institute of Computing Technology, Chinese Academy of Sciences

< P4D-08 >

Outlines

- Introduction
- Methodology
- Experiments
- Conclusion

Why Need Recommendation System?

- We are facing hundreds and thousands of options in the internet, having no clue about a decision.
- Recommendation System (RS) plays an important role in helping users find the most relevant and interesting new objects for them based on their historical behavior records.

User behaviors in Amazon.com

Share

New Apple MacBook Pro (15-inch, Touch Bar, 2.3GHz 8-core Intel Core i9, 16GB RAM, 512GB SS
Space Gray

57 customer reviews | 118 answered questions

Roll over image to zoom in

List Price: ~~$2,799.00~~
Price: $2,499.00
You Save: $300.00 (11%)

Note: Signature required upon delivery due to high value of this item. Details

In Stock.

This item does not ship to **Taiwan; Republic of China**. Please check other sellers who may ship internationally. Learn more
Ships from and sold by Amazon.com.

Style: **Intel Core I9**

Intel Core I7 | Intel Core I9

Capacity: **512GB**

256GB | 512GB

Color: **Space Gray**

- 9Th-generation 8-Core Intel Core i9 Processor
- Brilliant Retina Display with True Tone technology
- Touch Bar and Touch ID
- Radeon Pro 560x Graphics with 4GB of video Memory
- Ultrafast SSD
- Intel UHD Graphics 630
- Four Thunderbolt 3 (USB-C) ports

Show more

Jump to: Compare devices | Technical details

New (5) from $2,499.00 Details

Qty: 1

Add to Cart

Buy Now

Add gift options

Add Additional Items

- AppleCare+ for Mac (3 years) $379.00
- Apple Magic Keyboard with Numeric Keypad - Silver ~~$99.00~~ $88.99
- Apple Magic Trackpad 2 - Silver: Silver ~~$129.00~~ $118.50
- Apple 87W USB-C Power Adapter ~~$79.00~~ $69.00
- Save $20 at checkout | Microsoft Office 365 Home: Download ~~$99.99~~ $79.99

Deliver to Taiwan

Add to List

Other Sellers on Amazon

$2,549.92 Add to Cart
Sold by: Expercom - Apple Premier Partner

$2,569.00 Add to Cart
Sold by: Adorama Camera

Rich Behavior Records

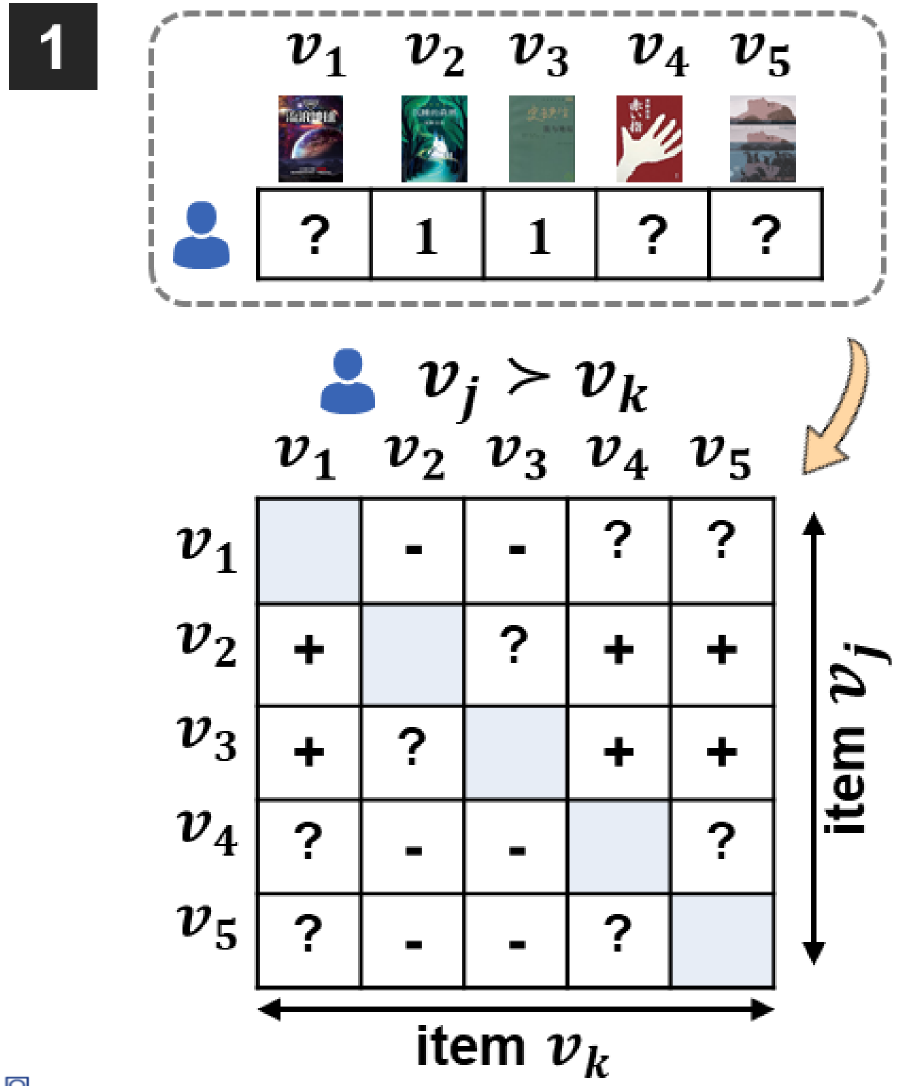
- Explicit Feedback
  - Directly reflect the preference of users toward objects
  - E.g. star ratings, thumbs up/down, like
  - It is not always available

- Implicit Feedback
  - Only positive feedback is available
  - E.g. purchase history, watching habits, mouse movements.
  - More abundant than explicit feedback

Motivations

- Traditional matrix factorization based methods does not satisfy the triangle inequality leading to sub-optimal performance.
- Most of the existing algorithms merely focus on datasets with sufficient amount of samples, with few considers how to avoid overfitting when confronting sparse and insufficient preference information.

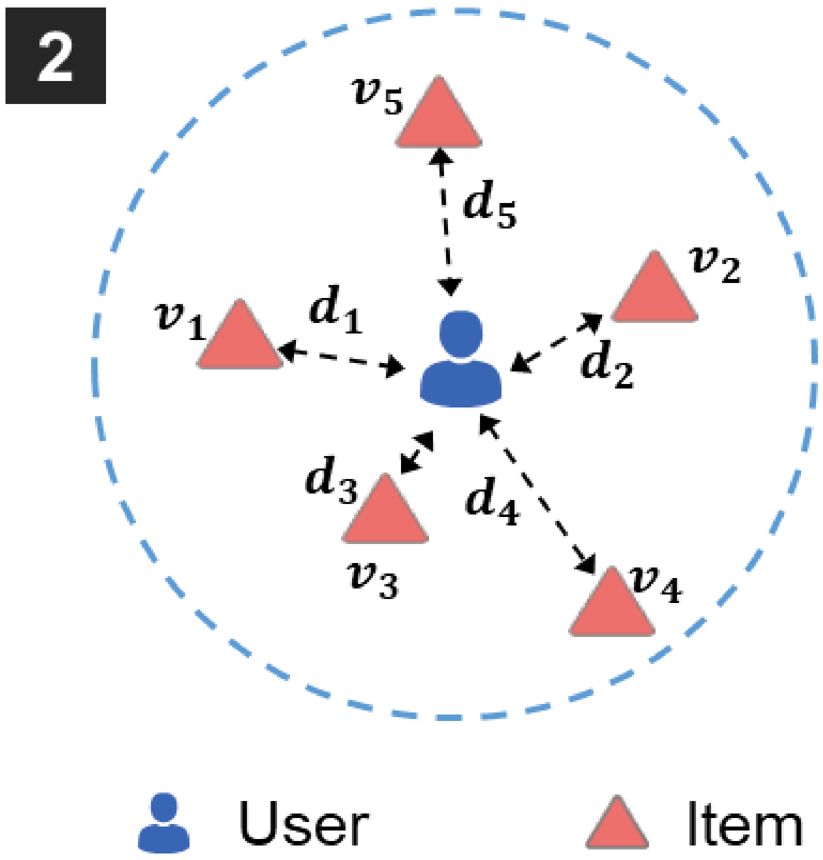
Collaborative Preference Embedding



- A set of users $\mathcal{U} = \{u_1, u_2, \dots, u_M\}$
- A set of items $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$
- For a triplet (i, j, k), if user u_i prefers v_j to v_k, we denote this relation as $v_j \succ_{u_i} v_k$. If u_i prefers v_k to v_j, denote as $v_j \prec_{u_i} v_k$.
- $\mathcal{Y} = \left\{y_{jk}^{(i)}\right\}$ is the set of triplet labels:

$$y_{jk}^{(i)} = \begin{cases} 1, & v_j \succ_{u_i} v_k \\ -1, & otherwise \end{cases}$$

Preserve Preference Consistency



$v_j \succ v_k$: user prefers v_j to v_k

- Learn an embedding space where we can capture the preference via comparing the relative Euclidean distances. We expect:

$$\begin{cases} \mathbf{d}(i,j) < \mathbf{d}(i,k), & v_j \succ_{u_i} v_k \\ \mathbf{d}(i,j) > \mathbf{d}(i,k), & v_j \prec_{u_i} v_k \end{cases}$$

- Margin function

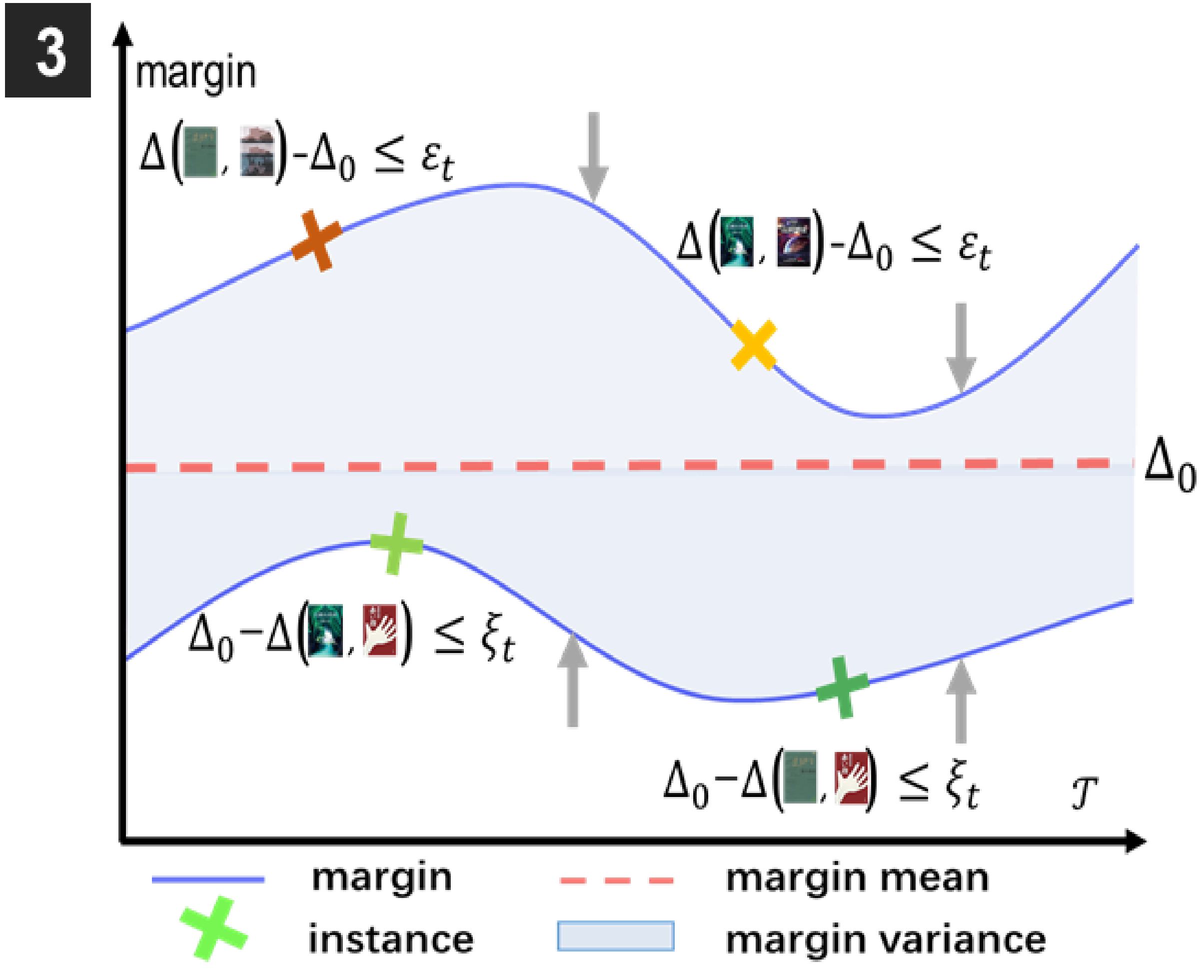
learned embedding space

$$\Delta_{jk}^{(i)} = y_{jk}^{(i)} \cdot \left(\mathbf{d}(i,k)^2 - \mathbf{d}(i,j)^2\right)$$
$$= \underbrace{y_{jk}^{(i)}}_{\text{true user preference}} \cdot \underbrace{\left(||\mathbf{f}_{u_i} - \mathbf{f}_{v_k}||^2 - ||\mathbf{f}_{u_i} - \mathbf{f}_{v_j}||^2\right)}_{\text{learned embedding space}}$$

true user preference

$f_{u_i} f_{v_j}$ are the learned embeddings

Optimize the Margin Distribution



- set the margin mean as a constant Δ_0:

$$\Delta_0 = \frac{1}{|\mathcal{T}|} \cdot \sum_{(i,j,k)\in\mathcal{T}} \Delta_{jk}^{(i)}$$

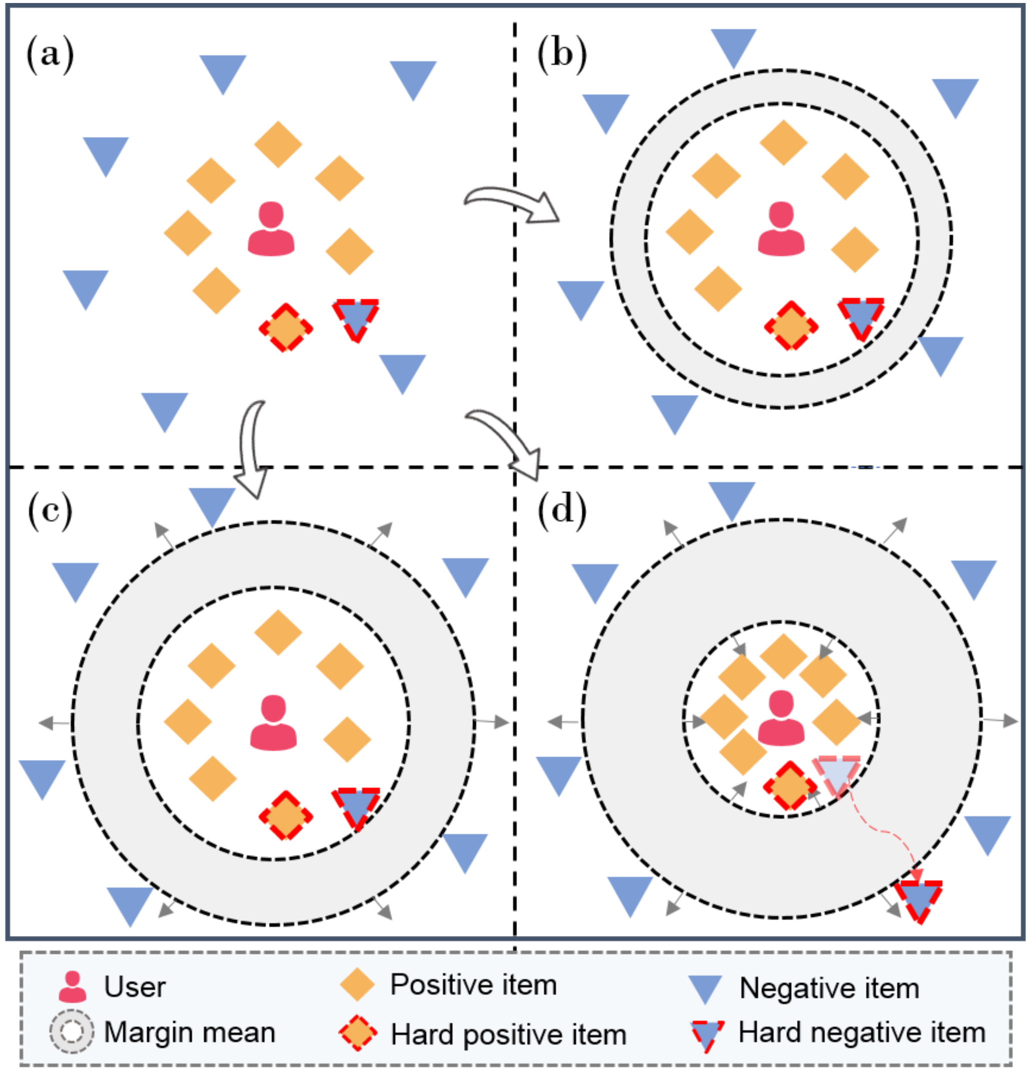
- restrict the margin deviation:

$$\left|\Delta_{jk}^{(i)} - \Delta_0\right|$$

- maximize the margin mean and minimize the margin variance:

$$\underset{\mathbf{f}_u,\mathbf{f}_v}{\mathbf{argmin}} \frac{1}{|\mathcal{T}|} \sum_{(i,j,k)\in\mathcal{T}} \mathbf{max}\left(\Delta_{jk}^{(i)} - \Delta_0, 0\right) + \mathbf{max}\left(\Delta_0 - \Delta_{jk}^{(i)}, 0\right)$$

Optimize the Margin Distribution



Why optimize margin distribution?

(a) Initial space
(b) The effects of small margin mean
(c) The effects of large margin mean and large margin variance
(d) The embedding space with large margin mean and small margin variance

Leverage a Compact Space

- Measure the correlation between different dimensions with the covariance matrix C:

the average embedding

$$C = \frac{1}{N+M}\sum_{i=1}^{N+M}(\mathbf{f}_i - \bar{\mathbf{f}})^\top(\mathbf{f}_i - \bar{\mathbf{f}})$$

- Adopt the log-determinant divergence (LDD) to reduce the redundancy between different dimensions

a diagonal matrix to avoid numerical disaster when C is a singular matrix

$$\tilde{\mathcal{R}}_C = tr(C) - \log\left(\det(C + \delta \cdot I)\right) = tr(C) - \sum_{i=1}^{d}\log(\lambda_i + \delta)$$

$$\log\left(det(C)\right) = \log\left(\prod_{i=1}^{d}\lambda_i\right)$$

the eigenvalues of C

The Overall Objective Function

loss function for optimizing the margin distribution

$$\underset{\mathbf{f}_u,\mathbf{f}_v,\xi,\epsilon}{\textbf{argmin}} \ \frac{1}{|\mathcal{T}|} \sum_{(i,j,k)\in\mathcal{T}} \textbf{max}\left(\Delta_{jk}^{(i)} - \Delta_0, 0\right) + \textbf{max}\left(\Delta_0 - \Delta_{jk}^{(i)}, 0\right)$$

$$+ \mu \cdot \left(tr(C) - \sum_{i=1}^{d} \log(\lambda_i + \delta) \right)$$

LDD loss

$$s.t. \ ||\mathbf{f}_{\boldsymbol{u}_i}||^2 \leq l, ||\mathbf{f}_{\boldsymbol{v}_j}||^2 \leq l$$

bounded L2 norm

Experiments

- We conduct extensive experiments on three benchmark datasets.

Datasets	MovieLens-100K	CiteULike-T	Book-Crossing
Domain	Movie	Paper	Book
#Users	943	7,947	11,209
#Items	1,682	25,975	7,490
#Ratings	55,376	142,794	98,205
%Density	3.4912%	0.0692%	0.1170%

- Note that, the smaller the value of %Density is, the more sparsely labeled the dataset is. Experiments on the last two datasets can evaluate the performance of CPE when facing with sparse and insufficient preference information.

Experiments

We evaluate CPE against 5 competitors and evaluate all methods with 5 widely used metrics.

Method	P@30↑	R@30↑	NDCG@30↑	P@50↑	R@50↑	NDCG@50↑	MAP↑	AUC↑
GMF[9]	0.1000	0.0449	0.1046	0.1560	0.0464	0.1645	0.0333	0.7302
NeuMF[9]	0.0917	0.0452	0.0983	0.1480	0.0521	0.1553	0.0328	0.7127
BPR-MF[27]	0.1794	0.1003	0.2041	0.2009	0.1130	0.2274	0.0929	0.8538
WRMF[13]	0.1681	0.0930	0.1943	0.2048	0.1083	0.2207	0.0864	0.8284
CML-PAIR[12]	0.1656	0.0832	0.2006	0.2110	0.1051	0.2560	0.0709	0.8173
CML-WARP[12]	0.1889	0.0955	0.2241	0.2311	0.1297	0.2851	0.0838	0.8474
CPE (ours)	0.2111	0.1118	0.2356	0.2525	0.1645	0.2961	0.1079	0.8699

(b) CiteULike-T

Method	P@30↑	R@30↑	NDCG@30↑	P@50↑	R@50↑	NDCG@50↑	MAP↑	AUC↑
GMF[9]	0.3107	0.2393	0.3527	0.3556	0.3001	0.4249	0.2468	0.8815
NeuMF[9]	0.3567	0.2731	0.4061	0.4116	0.3740	0.4778	0.3054	0.9053
BPR-MF[27]	0.3557	0.2714	0.4080	0.4050	0.3674	0.4700	0.2882	0.9027
WRMF[13]	0.3433	0.2662	0.3734	0.3800	0.3469	0.4129	0.2822	0.8890
CML-PAIR[12]	0.3384	0.2476	0.3941	0.3809	0.3047	0.4365	0.2566	0.8475
CML-WARP[12]	0.3402	0.2485	0.3949	0.3835	0.3076	0.4354	0.2616	0.8596
CPE (ours)	0.3633	0.2793	0.4002	0.4283	0.3910	0.4957	0.2938	0.9056

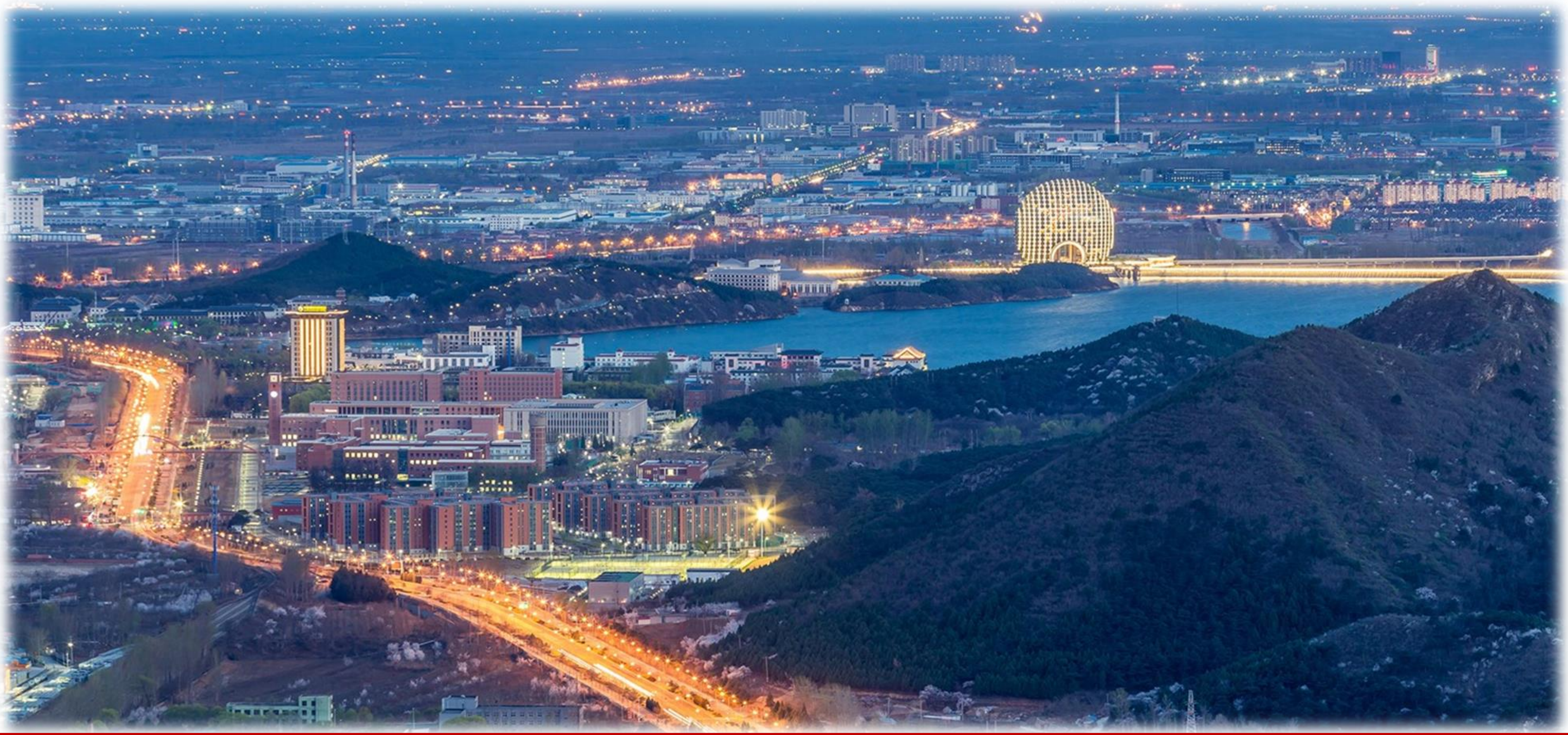
(a) MovieLens-100K

Method	P@30↑	R@30↑	NDCG@30↑	P@50↑	R@50↑	NDCG@50↑	MAP↑	AUC↑
GMF[9]	0.0833	0.0311	0.1092	0.1400	0.0412	0.1310	0.0421	0.6405
NeuMF[9]	0.0778	0.0312	0.0939	0.1501	0.0510	0.1618	0.0455	0.6385
BPR-MF[27]	0.1476	0.0792	0.1525	0.1833	0.1249	0.2092	0.0615	0.7278
WRMF[13]	0.1238	0.0681	0.1308	0.1767	0.1223	0.1929	0.0548	0.7109
CML-PAIR[12]	0.1734	0.0816	0.1890	0.2067	0.1525	0.2484	0.0561	0.7510
CML-WARP[12]	0.1810	0.1055	0.1975	0.2400	0.1782	0.2689	0.0836	0.7605
CPE (ours)	0.2067	0.1188	0.2161	0.2869	0.1952	0.3247	0.1038	0.8359

(c) Book-Crossing

Conclusion

- We develop a novel Collaborative Preference Embedding (CPE) to effectively address the problem of sparse and insufficient preference supervision in RS.
- To alleviate the limited generalization ability, we devise a margin function and propose a generalization enhancement scheme by optimizing the margin distribution.
- we adopt a novel regularization strategy to leverage a compact embedding space, which can further enhance the generalization performance.



THANKS !