# Collaborative Preference Embedding against Sparse Labels

ACMmultimedia
Nice, France 21-25 October 2019

Shilong Bao[1,2], Qianqian Xu[3], Ke Ma[1,2], Zhiyong Yang[1,2], Xiaochun Cao[1,2], Qingming Huang[2,3,4*]
[1]Institute of Information Engineering, CAS [2]University of Chinese Academy of Sciences
[3]Institute of Computing Technology, CAS [4]BDKM, CAS
{baoshilong, make, yangzhiyong, caoxiaochun}@iie.ac.cn, xuqianqian@ict.ac.cn, qmhuang@ucas.ac.cn

## Motivation

- Traditional MF-based methods fail to satisfy the triangle inequality and lead to sub-optimal performance.
- Furthermore, the vast majority of existing algorithms merely focus on datasets with sufficient amount of samples, which limits the generalization performance and easily leads to overfitting problems, when facing with sparse and insufficient preference information.

## Framework

- **Learn an embedding space** where we can infer the user preference by comparing the relative Euclidean distances between the user and different items. We expect the following relations:

$$\mathrm{d}(i,j) < \mathrm{d}(i,k), \quad v_j \succ_{u_i} v_k$$
$$\mathrm{d}(i,j) > \mathrm{d}(i,k), \quad v_j \prec_{u_i} v_k$$

where $v_j \succ_{u_i} v_k$ represents that user $u_i$ prefers $v_j$ to $v_k$.

- **Margin function** which could indicate the consistency between the embedding space and the true user preference.

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

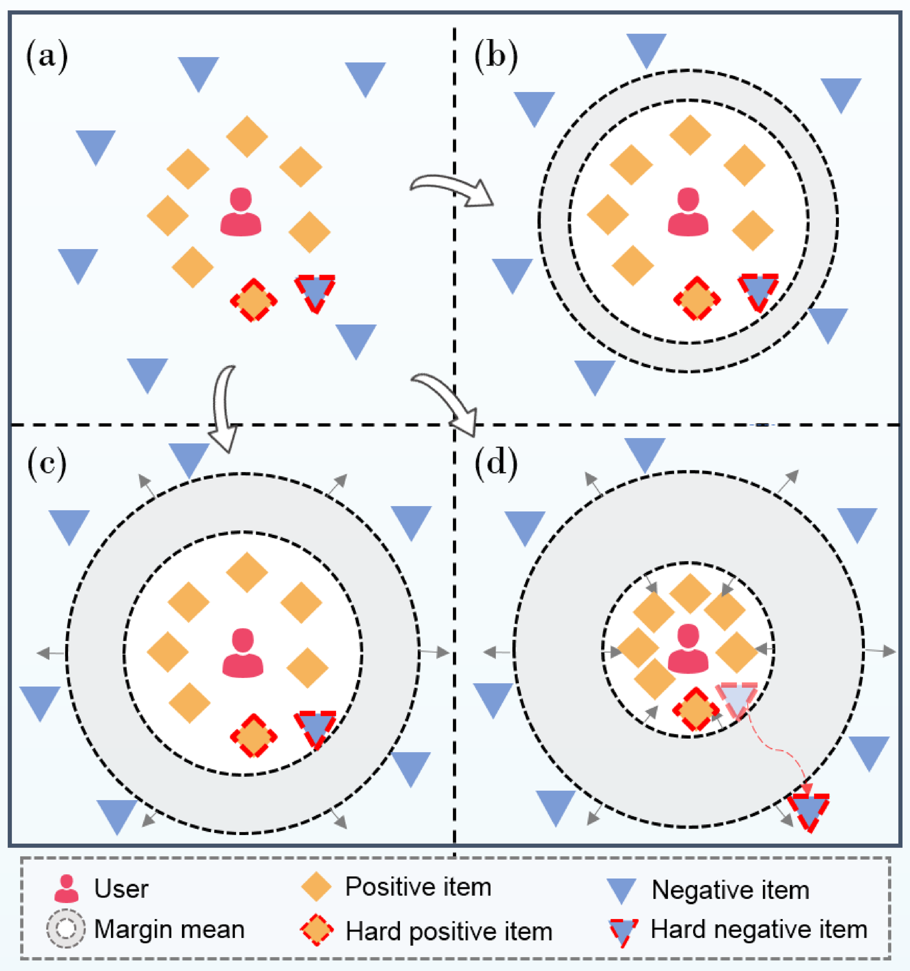
true user preference label

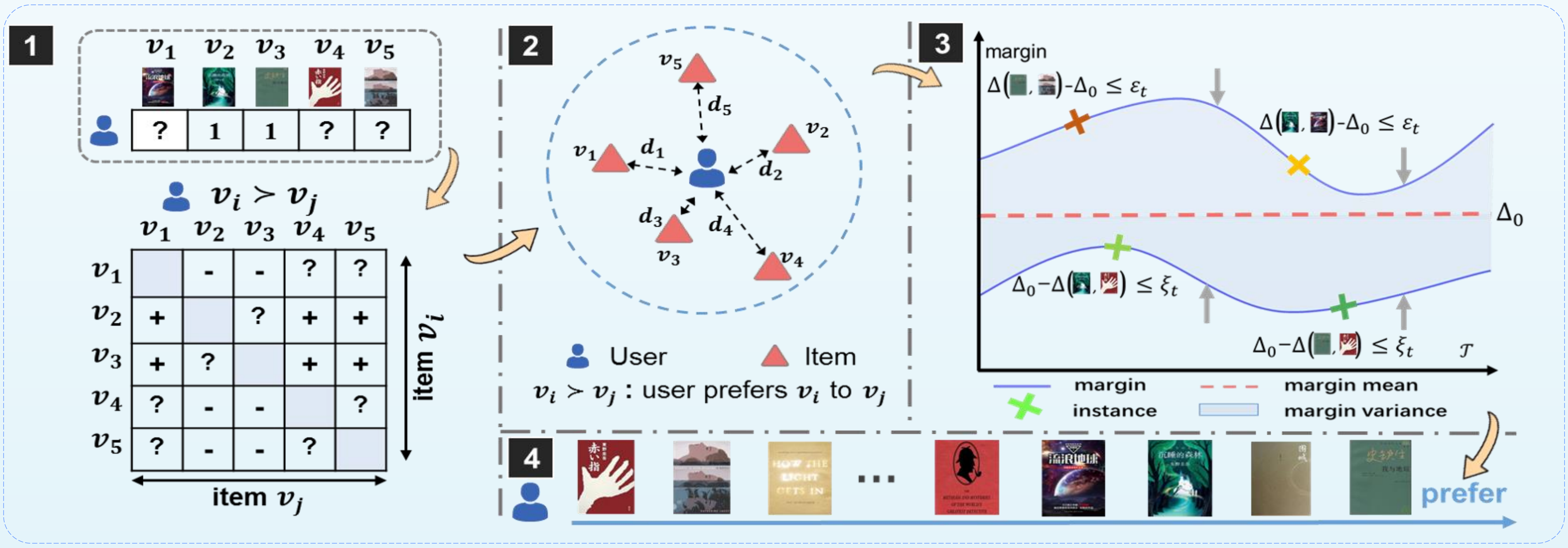
$$\Delta_{jk}^{(i)} = y_{jk}^{(i)} \cdot \left(\mathrm{d}(i,k)^2 - \mathrm{d}(i,j)^2\right)$$
$$= y_{jk}^{(i)} \cdot \left(||f_{u_i} - f_{v_k}||^2 - ||f_{u_i} - f_{v_j}||^2\right)$$

- **Optimize margin distribution**

An illustration of why optimizing the margin distribution of our metric space can achieve better generalization performance.

(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.** To reduce the redundancy between different dimensions, the **log-determinant divergence** (LDD) is adapted:

the eigenvalues of C

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

where $\delta \cdot I (\delta > 0)$ is a diagonal matrix to avoid numerical disaster when C is a singular matrix. $C$ is the covariance matrix to measure the correlation between different dimensions: the average embedding

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

- **Objective function**

the loss function for optimizing margin distribution

$$\mathrm{argmin}_{f_u, f_v, \xi, \epsilon} \ \frac{1}{|\mathcal{T}|} \sum_{(i,j,k) \in \mathcal{T}} \max\left(\Delta_{jk}^{(i)} - \Delta_0, 0\right) + \max\left(\Delta_0 - \Delta_{jk}^{(i)}, 0\right)$$
$$+ \mu \cdot \left(tr(C) - \sum_{i=1}^{d} \log(\lambda_i + \delta)\right) \quad \text{the LDD term}$$
$$s.t. \ \ ||f_{u_i}||^2 \le l, ||f_{v_j}||^2 \le l \quad \text{bounded L2 norm}$$

## Experiments

| 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] | 0.1000 | 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] | 0.0833 | 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**