Asymmetric Bayesian Personalized Ranking for One-Class Collaborative Filtering

Shan Ouyang, Lin Li, WeiKe Pan* and Zhong Ming*
{ouyangshan, lilin20171}@email.szu.edu.cn, {panweiike, mingz}@suzu.edu.cn

College of Computer Science and Software Engineering
Shenzhen University, Shenzhen, China
Introduction

Problem Definition

One-Class Collaborative Filtering

- **Input**: A set of (user, item) pairs $\mathcal{R} = \{(u, i)\}$, where each $(u, i)$ pair means that user $u$ has a positive feedback to item $i$.

- **Goal**: recommend each user $u \in \mathcal{U}$ a personalized ranked list of items from the set of unobserved items, i.e., $\mathcal{I} \setminus \mathcal{I}_u$. 

![Diagram showing one-class feedback matrix]
A user prefers an interacted item to an un-interacted item, e.g., user 3 prefers item 2 to item 4, i.e., \((3, 2) \succ (3, 4)\) or \(\hat{r}_{32} > \hat{r}_{34}\).

In general, we have,

\[ \hat{r}_{ui} > \hat{r}_{uj}, \ i \in \mathcal{I}_u, j \in \mathcal{I} \setminus \mathcal{I}_u. \]

Bayesian personalized ranking (BPR) [Rendle et al., 2009] is built on this assumption.
An item is preferred by an interacted user to an un-interacted user, e.g., item 2 is preferred by user 3 to user 6, i.e., $(3, 2) \succ (6, 2)$ or $\hat{r}_{32} > \hat{r}_{62}$.

In general, we have,

$$\hat{r}_{ui} > \hat{r}_{wi}, \ u \in U_i, \ w \in U \setminus U_i.$$  

We call a model built on this transposed preference assumption BPR$^\top$. 

---

**Vertical Pairwise Preference Assumption**
Combining those two types of pairwise preference assumptions, we have,

\[ \hat{r}_{ui} > \hat{r}_{uj}, \hat{r}_{ui} > \hat{r}_{wi}, i \in \mathcal{I}_u, j \in \mathcal{I} \setminus \mathcal{I}_u, u \in \mathcal{U}_i, w \in \mathcal{U} \setminus \mathcal{U}_i. \]

Mutual Bayesian personalized ranking (MBPR) [Yu et al., 2016] is built on this \textbf{symmetric} assumption.
Our Asymmetric Assumption

- The **symmetric** mutual pairwise preference assumption may not hold, in particular of the vertical one, because **different users may have different evaluation standards**, which will then make the preferences of different users uncomparable.

- We propose an **asymmetric** pairwise preference assumption
  - A user prefers an interacted item to an un-interacted item.
  - An item is preferred by a **group** of interacted users to a **group** of un-interacted users.
We propose a novel and improved preference assumption, i.e., asymmetric pairwise preference assumption, where we relax the vertical preference assumption to make it more reasonable and comparable.

With the proposed first asymmetric assumption for OCCF, we then design a novel recommendation algorithm called asymmetric Bayesian personalized ranking (ABPR).
### Table: Notations and descriptions.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>number of users</td>
</tr>
<tr>
<td>$m$</td>
<td>number of items</td>
</tr>
<tr>
<td>$u, w$</td>
<td>user ID</td>
</tr>
<tr>
<td>$i, j$</td>
<td>item ID</td>
</tr>
<tr>
<td>$r_{ui}$</td>
<td>preference of user $u$ to item $i$</td>
</tr>
<tr>
<td>$\mathcal{U} = {u}$</td>
<td>the whole set of users</td>
</tr>
<tr>
<td>$\mathcal{I} = {i}$</td>
<td>the whole set of items</td>
</tr>
<tr>
<td>$\mathcal{R} = {(u, i)}$</td>
<td>one-class feedback</td>
</tr>
<tr>
<td>$\mathcal{U}_i$</td>
<td>a set of users who interact with item $i$</td>
</tr>
<tr>
<td>$\mathcal{I}_u$</td>
<td>a set of items interacted by user $u$</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>a group of users who interact with an item</td>
</tr>
<tr>
<td>$\mathcal{N}$</td>
<td>a group of users who do not interact with an item</td>
</tr>
<tr>
<td>$\hat{r}_{ui}$</td>
<td>predicted preference of user $u$ to item $i$</td>
</tr>
<tr>
<td>$d$</td>
<td>number of latent dimensions</td>
</tr>
<tr>
<td>$U_u \in \mathbb{R}^{1 \times d}$</td>
<td>user $u$'s latent feature vector</td>
</tr>
<tr>
<td>$V_i \in \mathbb{R}^{1 \times d}$</td>
<td>item $i$'s latent feature vector</td>
</tr>
<tr>
<td>$b_i \in \mathbb{R}$</td>
<td>item $i$'s bias</td>
</tr>
</tbody>
</table>
Asymmetric Pairwise Preference Assumption (1/2)

We keep the horizontal pairwise preference assumption in BPR [Rendle et al., 2009] and assume that an item is preferred by a group of interacted users to a group of un-interacted users in order to make the vertical one more reasonable and comparable,

\[ \hat{r}_{ui} > \hat{r}_{uj}, \hat{r}_{P_i} > \hat{r}_{N_i}, \] (1)

where \( i \in I_u, j \in I \setminus I_u, P \subseteq U_i, N \subseteq U \setminus U_i \) and \( u \in P \).
For instantiation of the relationship between the group preferences $\hat{r}_{Pi}$ and $\hat{r}_{Ni}$, we propose “Many ‘Group vs. One’ (MGO)” inspired by “Many ‘Set vs. One’ (MSO)” [Pan et al., 2019],

$$\hat{r}_{Pi} > \hat{r}_{wi}, w \in N,$$

(2)

where $\hat{r}_{Pi} = \frac{1}{|P|} \sum_{u' \in P} \hat{r}_{u'i}$ is the overall preference of user-group $P$ to item $i$.

Notice that $\hat{r}_{u'i} = U_{u'} \cdot V_{i.}^T + b_{u'} + b_i$ is the prediction rule for the preference of user $u'$ to item $i$, where $U_{u'} \in \mathbb{R}^{1 \times d}$ and $V_{i.} \in \mathbb{R}^{1 \times d}$ are latent feature vectors of user $u'$ and item $i$, respectively, and $b_{u'} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are the bias of user $u'$ and item $i$, respectively.
Asymmetric Bayesian Personalized Ranking (ABPR)

Based on the asymmetric pairwise preference assumption in Eqs. (1-2), we reach an objective function in our asymmetric Bayesian personalized ranking (ABPR) for each quintuple \((u, i, j, P, N)\),

\[
\min_{\Theta} - \ln \sigma(\hat{r}_{uij}) - \frac{1}{|N|} \sum_{w \in N} \ln \sigma(\hat{r}_{iPw}) + \text{reg}(u, i, j, P, N),
\]

where \(\Theta = \{U_u, V_i, b_u, b_i, u \in U, i \in I\}\) are the model parameters to be learned, \(\hat{r}_{uij} = \hat{r}_{ui} - \hat{r}_{uj}\) and \(\hat{r}_{iPw} = \hat{r}_{P_i} - \hat{r}_{wi}\) denote the corresponding preference differences, and

\[
\text{reg}(u, i, j, P, N) = \frac{\alpha}{2} \| V_i \|^2 + \frac{\alpha}{2} \| V_j \|^2 + \frac{\alpha}{2} \| b_i \|^2 + \frac{\alpha}{2} \| b_j \|^2 + \sum_{u' \in P} \left[ \frac{\alpha}{2} \| U_{u'} \|^2 + \frac{\alpha}{2} \| b_{u'} \|^2 \right] + \sum_{w \in N} \left[ \frac{\alpha}{2} \| U_w \|^2 + \frac{\alpha}{2} \| b_w \|^2 \right]
\]

is the regularization term used to avoid overfitting.
Gradients

We then have the gradients of the model parameters w.r.t. the tentative objective function in Eq.(3),

\[
\nabla U_u = -\sigma(-\hat{r}_{uij})(V_i - V_j) - \sum_{w \in \mathcal{N}} \frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw}) \frac{V_i}{|\mathcal{P}|} + \alpha U_u, \\
\nabla U_{u'} = -\sum_{w \in \mathcal{N}} \frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw}) \frac{V_i}{|\mathcal{P}|} + \alpha U_{u'}, u' \in \mathcal{P}\{u\}, \\
\nabla U_w = -\frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw})(-V_i) + \alpha U_w, w \in \mathcal{N}, \\
\nabla V_i = -\sigma(-\hat{r}_{uij})U_u - \sum_{w \in \mathcal{N}} \frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw})(\sum_{u \in \mathcal{P}} \frac{U_u}{|\mathcal{P}|} - U_w) + \alpha V_i, \\
\nabla V_j = -\sigma(-\hat{r}_{uij})(-U_u) + \alpha V_j, \\
\n\nabla b_i = -\sigma(-\hat{r}_{uij}) + \alpha b_i, \\
\n\nabla b_j = -\sigma(-\hat{r}_{uij})(-1) + \alpha b_j, \\
\n\nabla b_u = -\sum_{w \in \mathcal{N}} \frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw}) \frac{1}{|\mathcal{P}|} + \alpha b_u, u \in \mathcal{P}, \\
\n\nabla b_w = -\frac{1}{|\mathcal{N}|} \sigma(-\hat{r}_{ipw})(-1) + \alpha b_w, w \in \mathcal{N}.
\]
## Method

**Algorithm**

1: for $t = 1, 2, \ldots, T$ do 
2:  \hspace{1em} for $t_2 = 1, 2, \ldots, |\mathcal{R}|$ do 
3:  \hspace{2em} Randomly pick a (user, item) pair $(u, i)$ from $\mathcal{R}$. 
4:  \hspace{2em} Randomly pick an item $j$ from $\mathcal{I} \setminus \mathcal{I}_u$. 
5:  \hspace{2em} Randomly pick $|\mathcal{P}| - 1$ users from $\mathcal{U}_i \setminus \{u\}$. 
6:  \hspace{2em} Randomly pick $|\mathcal{N}|$ users from $\mathcal{U} \setminus \mathcal{U}_i$. 
7:  \hspace{2em} Calculate the gradients w.r.t. the tentative objection function in Eq.(3). 
8:  \hspace{2em} Update the corresponding model parameters, i.e., $U_u, U_w, V_i, V_j, b_i, b_j, b_u,$ and $b_w$, where $u \in \mathcal{P}$ and $w \in \mathcal{N}$. 
9:  \hspace{2em} end for 
10: end for
Datasets

Table: Statistics of the first copy of each dataset used in the experiments. Notice that $n$ is the number of users and $m$ is the number of items, and $|\mathcal{R}|$, $|\mathcal{R}^{va}|$ and $|\mathcal{R}^{te}|$ denote the numbers of (user, item) pairs in training data, validation data and test data, respectively.

| Dataset       | $n$    | $m$    | $|\mathcal{R}|$ | $|\mathcal{R}^{va}|$ | $|\mathcal{R}^{te}|$ |
|---------------|--------|--------|-----------------|----------------------|----------------------|
| ML20M         | 138,493| 27,278 | 5,997,245       | 1,999,288            | 1,998,877            |
| NF50KU        | 50,000 | 17,770 | 3,551,369       | 1,183,805            | 1,183,466            |
Evaluation Metrics

We adopt five ranking-oriented metrics [Chen and Karger, 2006, Manning et al., 2008, Valcarce et al., 2018] to evaluate the performance:

- Precision@5
- Recall@5
- F1@5
- NDCG@5
- 1-call@5
In order to study the effectiveness of our proposed asymmetric pairwise preference assumption and the corresponding recommendation algorithm ABPR directly, we include the following closely related baseline methods, including

- (i) basic matrix factorization with square loss (MF),
- (ii) matrix factorization with logistic loss (LogMF) [Johnson, 2014],
- (iii) factored item similarity model (FISM) [Kabbur et al., 2013],
- (iv) Bayesian personalized ranking (BPR) [Rendle et al., 2009],
- (v) BPR with transposed pairwise preference assumption (BPR$^T$), and
- (vi) mutual BPR (MBPR) [Yu et al., 2016].

Notice that MF, LogMF and FISM are based on the pointwise preference assumption, and BPR, BPR$^T$ and MBPR are based on the pairwise preference assumptions.
Experiments

Parameter Configurations

For all the baseline methods and our ABPR, we implement them in the same SGD-based algorithmic framework written in Java for fair comparison\(^1\). In particular, we fix the number of latent dimensions \(d = 20\), the learning rate \(\gamma = 0.01\), and search the best value of the iteration number \(T \in \{10, 20, 30, \ldots, 990, 1000\}\) and the best value of the tradeoff parameter on the regularization terms \(\alpha \in \{0.001, 0.01, 0.1\}\) for each method on each dataset via the performance of NDCG@5 on the validation data. For MF, LogMF and FISM, we randomly sample three times of un-interacted (user, item) pairs as negative one-class feedback to augment the interacted (user, item) pairs for preference learning [Kabbur et al., 2013]. For our ABPR, we fix the number of user-group as \(|\mathcal{P}| = |\mathcal{N}| = 3\) [Pan et al., 2019].

---

\(^1\)The source code is available at http://csse.szu.edu.cn/staff/panwk/publications/ABPR/.

Ouyang et al. (SZU)
### Table: Recommendation performance of our ABPR, three pointwise preference learning methods, i.e., MF, LogMF and FISM, and three pairwise preference learning methods, i.e., BPR, BPR$^T$ and MBPR, on ML20M and NF50KU w.r.t. five commonly used ranking-oriented evaluation metrics. The significantly best results are marked in bold ($p$-value < 0.015).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Precision@5</th>
<th>Recall@5</th>
<th>F1@5</th>
<th>NDCG@5</th>
<th>1-call@5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ML20M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pointwise</td>
<td>MF</td>
<td>0.1249±0.0014</td>
<td>0.0755±0.0015</td>
<td>0.0773±0.0012</td>
<td>0.1378±0.0014</td>
<td>0.4479±0.0028</td>
</tr>
<tr>
<td></td>
<td>LogMF</td>
<td>0.1622±0.0002</td>
<td>0.0918±0.0002</td>
<td>0.0951±0.0001</td>
<td>0.1805±0.0001</td>
<td>0.5269±0.0007</td>
</tr>
<tr>
<td></td>
<td>FISM</td>
<td>0.1351±0.0014</td>
<td>0.0821±0.0010</td>
<td>0.0836±0.0009</td>
<td>0.1505±0.0019</td>
<td>0.4755±0.0028</td>
</tr>
<tr>
<td>Pairwise</td>
<td>BPR</td>
<td>0.1645±0.0009</td>
<td>0.0862±0.0007</td>
<td>0.0921±0.0007</td>
<td>0.1810±0.0013</td>
<td>0.5228±0.0016</td>
</tr>
<tr>
<td></td>
<td>BPR$^T$</td>
<td>0.0632±0.0024</td>
<td>0.0364±0.0015</td>
<td>0.0373±0.0015</td>
<td>0.0693±0.0025</td>
<td>0.2536±0.0078</td>
</tr>
<tr>
<td></td>
<td>MBPR</td>
<td>0.1609±0.0004</td>
<td>0.0893±0.0007</td>
<td>0.0931±0.0006</td>
<td>0.1797±0.0004</td>
<td>0.5262±0.0025</td>
</tr>
<tr>
<td></td>
<td>ABPR</td>
<td>0.1709±0.0005</td>
<td>0.0940±0.0005</td>
<td>0.0985±0.0004</td>
<td>0.1907±0.0007</td>
<td>0.5482±0.0016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NF50KU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pointwise</td>
<td>MF</td>
<td>0.1385±0.0002</td>
<td>0.0497±0.0009</td>
<td>0.0581±0.0005</td>
<td>0.1464±0.0008</td>
<td>0.4685±0.0019</td>
</tr>
<tr>
<td></td>
<td>LogMF</td>
<td>0.1665±0.0012</td>
<td>0.0580±0.0005</td>
<td>0.0671±0.0004</td>
<td>0.1773±0.0007</td>
<td>0.5185±0.0016</td>
</tr>
<tr>
<td></td>
<td>FISM</td>
<td>0.1447±0.0017</td>
<td>0.0520±0.0010</td>
<td>0.0606±0.0008</td>
<td>0.1536±0.0024</td>
<td>0.4833±0.0028</td>
</tr>
<tr>
<td>Pairwise</td>
<td>BPR</td>
<td>0.1664±0.0015</td>
<td>0.0560±0.0006</td>
<td>0.0654±0.0006</td>
<td>0.1765±0.0011</td>
<td>0.5112±0.0025</td>
</tr>
<tr>
<td></td>
<td>BPR$^T$</td>
<td>0.0912±0.0048</td>
<td>0.0302±0.0014</td>
<td>0.0349±0.0017</td>
<td>0.0971±0.0057</td>
<td>0.3275±0.0139</td>
</tr>
<tr>
<td></td>
<td>MBPR</td>
<td>0.1794±0.0012</td>
<td>0.0659±0.0009</td>
<td>0.0748±0.0008</td>
<td>0.1928±0.0013</td>
<td>0.5489±0.0033</td>
</tr>
<tr>
<td></td>
<td>ABPR</td>
<td>0.1854±0.0008</td>
<td>0.0685±0.0004</td>
<td>0.0777±0.0004</td>
<td>0.1995±0.0004</td>
<td>0.5601±0.0010</td>
</tr>
</tbody>
</table>
(i) our ABPR performs significantly better ($p$-value is smaller than 0.015) than all the baseline methods in all cases, which clearly shows the effectiveness of our proposed asymmetric pairwise preference assumption in modeling one-class feedback;

(ii) MBPR performs similar to BPR on ML20M and better than BPR on NF50KU, which shows the sensitivity of the mutual pairwise preference assumption in MBPR w.r.t. different datasets (notice that our ABPR performs better than BPR on both datasets showcasing the superiority of our relaxed vertical pairwise relationship in our asymmetric assumption);
(iii) the performance of BPR is much better than that of BPR$^T$, which is expected as the horizontal preference relationship, i.e., $(u, i) \succ (u, j)$, is more reasonable than the transposed one, i.e., $(u, i) \succ (w, i)$, considering the probable incomparability of preferences between different users (notice that our ABPR with asymmetric assumption, i.e., $(u, i) \succ (u, j)$ and $(P, i) \succ (N, i)$, performs the best); and

(iv) for the methods with pointwise preference assumption, LogMF performs similar to BPR and better than MF and FISM, which shows the importance of an appropriate loss function in modeling one-class feedback.
Related Work

1. Pointwise preference assumption
   - matrix factorization with square loss [Pan et al., 2008, Hu et al., 2008, He et al., 2016]
   - matrix factorization with logistic loss (LogMF) [Johnson, 2014]
   - factored item similarity model (FISM) [Kabbur et al., 2013]
   - deep learning based methods [Wu et al., 2016, He et al., 2017]

2. Pairwise preference assumption
   - Bayesian personalized ranking (BPR) [Rendle et al., 2009]
   - Collaborative filtering via learning pairwise preference over item-sets (CoFiSet) [Pan and Chen, 2013, Pan et al., 2019]
   - mutual BPR (MBPR) [Yu et al., 2016]
Conclusions

- We study an important recommendation problem called one-class collaborative filtering (OCCF) with users’ one-class feedback such as “likes” in many online and mobile applications.

- We propose a novel preference assumption called asymmetric pairwise preference assumption, where we assume that a user prefers an interacted item to an un-interacted one as well as an interacted item is preferred by a group of interacted users to a group of un-interacted users.

- We then design a novel recommendation algorithm called asymmetric Bayesian personalized ranking (ABPR), and find it performs significantly better than several pointwise preference learning methods and pairwise preference learning methods on two large and public datasets.
Future Work

For future works, we are interested in studying the proposed asymmetric preference assumption in more learning paradigms and problem settings such as listwise preference learning [Wu et al., 2018], deep learning [He et al., 2017, Lian et al., 2018], and sparsity reduction and cold-start recommendation [Lee et al., 2018].
We thank the support of National Natural Science Foundation of China Nos. 61872249, 61836005 and 61672358.
Less is more: Probabilistic models for retrieving fewer relevant documents.

Neural collaborative filtering.

Fast matrix factorization for online recommendation with implicit feedback.

Collaborative filtering for implicit feedback datasets.
In *Proceedings of the 8th IEEE International Conference on Data Mining*, ICDM’08, pages 263–272.

Logistic matrix factorization for implicit feedback data.
In *Proceedings of the NeurIPS 2014 Workshop on Distributed Machine Learning and Matrix Computations*.

FISM: Factored item similarity models for Top-N recommender systems.
In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD’13, pages 659–667.

gOCCF: Graph-theoretic one-class collaborative filtering based on uninteresting items.
In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, AAAI’18, pages 3448–3456.

xDeepFM: Combining explicit and implicit feature interactions for recommender systems.
In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD’18, pages 1754–1763.

*Introduction to Information Retrieval*.
Cambridge University Press, New York, NY, USA.

One-class collaborative filtering.
In *Proceedings of the 8th IEEE International Conference on Data Mining*, ICDM’08, pages 502–511.

CoFiSet: Collaborative filtering via learning pairwise preferences over item-sets.
In *Proceedings of SIAM International Conference on Data Mining*, SDM’13, pages 180–188.

Personalized recommendation with implicit feedback via learning pairwise preferences over item-sets.
*Knowledge and Information Systems*, 58(2):295–318.

BPR: Bayesian personalized ranking from implicit feedback.

On the robustness and discriminative power of information retrieval metrics for top-n recommendation.

SQL-Rank: A listwise approach to collaborative ranking.

Collaborative denoising auto-encoders for top-n recommender systems.