Transfer Learning for Heterogeneous One-Class Collaborative Filtering

Weike Pan, Mengsi Liu and Zhong Ming*

panweike@szu.edu.cn, liumengsi@email.szu.edu.cn, mingz@szu.edu.cn
College of Computer Science and Software Engineering
Shenzhen University
For a user $u \in \mathcal{U}$, we have a set of preferred items, i.e., $\mathcal{P}_u$, and a set of examined items, i.e., $\mathcal{E}_u$. Our goal is then to exploit such two types of one-class feedbacks and recommend a ranked list of items from $\mathcal{I} \setminus \mathcal{P}_u$ for each user $u$. 
Challenges

- Sparsity of positive feedbacks
- Uncertainty of implicit examinations
Algorithm: *Transfer via Joint Similarity Learning (TJSL)*
- Learn a similarity between a candidate item and a preferred item
- Learn a similarity between a candidate item and an identified likely-to-prefer examined item
Joint similarity learning has the merit of being able to connect two seemingly not related items w.r.t. the sparse positive feedbacks only.
### Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{U}$</td>
<td>user set, $u \in \mathcal{U}$</td>
</tr>
<tr>
<td>$\mathcal{I}$</td>
<td>item set, $i, i', j \in \mathcal{I}$</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>positive feedbacks</td>
</tr>
<tr>
<td>$\mathcal{P}_u = {i</td>
<td>(u, i) \in \mathcal{P}}$</td>
</tr>
<tr>
<td>$\mathcal{E}$</td>
<td>implicit examinations</td>
</tr>
<tr>
<td>$\mathcal{E}_u = {i</td>
<td>(u, i) \in \mathcal{E}}$</td>
</tr>
<tr>
<td>$\mathcal{R} = {(u, i)}$</td>
<td>all (user, item) pairs</td>
</tr>
<tr>
<td>$\mathcal{A} \subset \mathcal{R} \setminus \mathcal{P}$</td>
<td>sampled feedbacks</td>
</tr>
<tr>
<td>$\mathcal{P}^{te}$</td>
<td>test positive feedbacks</td>
</tr>
<tr>
<td>$d \in \mathbb{R}$</td>
<td>latent feature number</td>
</tr>
<tr>
<td>$V_i, P_i, E_j \in \mathbb{R}^{1 \times d}$</td>
<td>latent feature vectors</td>
</tr>
<tr>
<td>$b_u \in \mathbb{R}$</td>
<td>user bias</td>
</tr>
<tr>
<td>$b_i \in \mathbb{R}$</td>
<td>item bias</td>
</tr>
<tr>
<td>$r_{ui} \in {1, 0}$</td>
<td>preference of $(u, i)$</td>
</tr>
<tr>
<td>$\hat{r}_{ui}$</td>
<td>prediction of $(u, i)$</td>
</tr>
<tr>
<td>$T, L, L_0$</td>
<td>iteration number</td>
</tr>
<tr>
<td>$\rho$</td>
<td>sampling parameter</td>
</tr>
<tr>
<td>$\alpha_v, \alpha_p, \alpha_e, \beta_u, \beta_v$</td>
<td>tradeoff parameters</td>
</tr>
</tbody>
</table>
The predicted preference of user $u$ on item $i$, 

$$\hat{r}_{ui} = b_u + b_i + \sum_{i' \in P_u \setminus \{i\}} s_{i'i}, \quad (1)$$

where $\sum_{i' \in P_u \setminus \{i\}} s_{i'i} = \overline{U}_{u.}^{-i} V_{i.}^T$, and 

$$\overline{U}_{u.}^{-i} = \frac{1}{\sqrt{|P_u \setminus \{i\}|}} \sum_{i' \in P_u \setminus \{i\}} P_{i'..}.$$

Note that $\overline{U}_{u.}^{-i}$ can be considered as a virtual user profile w.r.t. user $u$ and item $i$ according to positive feedbacks.
Prediction Rule of TJSL(1,1) for HOCCF

The predicted preference of user $u$ on item $i$,

$$
\hat{r}_{ui} = b_u + b_i + \sum_{i' \in \mathcal{P}_u \setminus \{i\}} s_{i'i} + \sum_{j \in \mathcal{E}_u} s_{ji},
$$

(2)

where $\sum_{i' \in \mathcal{P}_u \setminus \{i\}} s_{i'i} = \bar{U}_{u,i} V_i^T$, $\sum_{j \in \mathcal{E}_u} s_{ji} = \tilde{U}_{u,i} V_i^T$, and

$$
\bar{U}_{u,i} = \frac{1}{\sqrt{|\mathcal{P}_u \setminus \{i\}|}} \sum_{i' \in \mathcal{P}_u \setminus \{i\}} P_{i'i},
$$

$$
\tilde{U}_{u,i} = \frac{1}{\sqrt{|\mathcal{E}_u|}} \sum_{j \in \mathcal{E}_u} E_{j,i}.
$$

Note that $\tilde{U}_{u,i}$ is a virtual user profile of user $u$ according to implicit examinations.
Objective Function of TJSL(1,1)

The objective function of TJSL(1,1),

$$\min_{\Theta} \sum_{(u,i) \in P \cup A} f_{ui},$$  \hspace{1cm} (3)$$

where $\Theta = \{b_u, b_i, V_i, P_{i'}, E_j\}$, $i, i', j = 1, \ldots, m$, $u = 1, \ldots, n$, and

$$f_{ui} = \frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + \frac{\alpha_v}{2} \|V_i\|_F^2 + \frac{\alpha_p}{2} \sum_{i' \in P\setminus\{i\}} \|P_{i'}\|_F^2 + \frac{\alpha_e}{2} \sum_{j \in E_u} \|E_j\|_F^2 + \frac{\beta_u}{2} b_u^2 + \frac{\beta_v}{2} b_i^2.$$
Gradients of TJSL(1,1)

For each \((u, i) \in P \cup A\), we have the gradients,

\[
\nabla b_u = \frac{\partial f_{ui}}{\partial b_u} = -e_{ui} + \beta_u b_u
\]

\[
\nabla b_i = \frac{\partial f_{ui}}{\partial b_i} = -e_{ui} + \beta_v b_i
\]

\[
\nabla V_i. = \frac{\partial f_{ui}}{\partial V_i.} = -e_{ui}(\bar{U}_{u.i} - i + \tilde{U}_u.) + \alpha_v V_i.
\]

\[
\nabla P_{i'.} = \frac{\partial f_{ui}}{\partial P_{i'.}} = -e_{ui}\frac{1}{\sqrt{|P_u \setminus \{i\}|}} V_i. + \alpha_p P_{i'}, i' \in P_u \setminus \{i\}
\]

\[
\nabla E_j. = \frac{\partial f_{ui}}{\partial E_j.} = -e_{ui}\frac{1}{\sqrt{|E_u|}} V_i. + \alpha_e E_{j}, j \in E_u
\]

where \(e_{ui} = r_{ui} - \hat{r}_{ui}\). Note that \(r_{ui} = 1\) if \((u, i) \in P\), and \(r_{ui} = 0\) if \((u, i) \in A\).
Update Rules of TJSL(1,1)

For each \((u, i) \in \mathcal{P} \cup \mathcal{A}\), we have the update rules,

\[
\begin{align*}
    b_u &= b_u - \gamma \nabla b_u \\
    b_i &= b_i - \gamma \nabla b_i \\
    V_i. &= V_i. - \gamma \nabla V_i. \\
    P_{i'} &= P_{i'} - \gamma \nabla P_{i'}, \quad i' \in \mathcal{P}_u \setminus \{i\} \\
    E_j. &= E_j. - \gamma \nabla E_j., \quad j \in \mathcal{E}_u
\end{align*}
\]

where \(e_{ui} = r_{ui} - \hat{r}_{ui}\).
Method

Algorithm of TJSL(1,1)

1: **Input**: Positive feedbacks $\mathcal{P}$, implicit examinations $\mathcal{E}$, iteration number $T$, and parameters $\rho, \alpha_v, \alpha_p, \alpha_e, \beta_u, \beta_v$.

2: **Output**: Learned model $\Theta$

3: Initialize the model $\Theta$

4: **for** $t = 1, \ldots, T$ **do**

5: Randomly sample $\mathcal{A} \subset \mathcal{R} \setminus \mathcal{P}$ with $|\mathcal{A}| = \rho |\mathcal{P}|$

6: **for** $t_2 = 1, \ldots, |\mathcal{P} \cup \mathcal{A}|$ **do**

7: Randomly pick up $(u, i) \in \mathcal{P} \cup \mathcal{A}$

8: Calculate $\bar{U}_{u}^{-i} = \frac{1}{\sqrt{|\mathcal{P}_u \setminus \{i\}|}} \sum_{i' \in \mathcal{P}_u \setminus \{i\}} P_{i'}$

9: Calculate $\tilde{U}_u = \frac{1}{\sqrt{|\mathcal{E}_u|}} \sum_{j \in \mathcal{E}_u} E_j$

10: Calculate $\hat{r}_{ui} = b_u + b_i + \bar{U}_{u}^{-i} V_{i}^T + \tilde{U}_u \cdot V_{i}^T$

11: Calculate $e_{ui} = r_{ui} - \hat{r}_{ui}$

12: Update $b_u, b_i, V_i, P_{i'}, i' \in \mathcal{P}_u \setminus \{i\}$ and $E_j, j \in \mathcal{E}_u$

13: **end for**

14: **end for**
Prediction Rule of TJSL

The predicted preference of user $u$ on item $i$,

$$\hat{r}_{ui}^{(\ell)} = b_u + b_i \sum_{i' \in \mathcal{P}_u \setminus \{i\}} s_{i'i} + \sum_{j \in \mathcal{E}_u^{(\ell)}} s_{ji}, \quad \mathcal{E}_u^{(\ell)} \subseteq \mathcal{E}_u,$$

(4)

where $s_{i'i} = \frac{1}{\sqrt{|\mathcal{P}_u \setminus \{i\}|}} P_{i'} V_i^T$, and $s_{ji} = \frac{1}{\sqrt{|\mathcal{E}_u^{(\ell)}|}} E_j V_i^T$. 
Objective Function of TJSL

The objective function of TJSL is as follows,

\[
\min_{\Theta^{(\ell)}, E^{(\ell)}_u \subseteq E_u \ (u,i) \in \mathcal{P} \cup \mathcal{A}} \sum_{u,i} f_{ui}^{(\ell)}
\]

where

\[
f_{ui}^{(\ell)} = \frac{1}{2}(r_{ui} - \hat{r}_{ui})^2 + \frac{\alpha_v}{2} \| V_i \|_F^2 + \frac{\alpha_p}{2} \sum_{i' \in \mathcal{P}_u \setminus \{i\}} \| P_{i'} \|_F^2 + \frac{\alpha_e}{2} \sum_{j \in E^{(\ell)}_u} \| E_j \|_F^2 + \frac{\beta_u}{2} b_u^2 + \frac{\beta_v}{2} b_i^2,
\]
and the model parameters are

\[
\Theta^{(\ell)} = \{ b_u, b_i, V_i., P_{i'}, E_j. | u \in \mathcal{U}, i \in \mathcal{I}, i' \in \mathcal{P}_u, j \in E^{(\ell)}_u \}.
\]
Gradients of TJSL

Given some selected examined items $\mathcal{E}_u^{(\ell)}$, we have the gradient of each corresponding $\theta \in \Theta^{(\ell)}$ for $(u, i) \in \mathcal{P} \cup \mathcal{A}$,

$$\nabla \theta = \frac{\partial f_{ui}^{(\ell)}}{\partial \theta}, \quad (6)$$

where $\nabla \theta$ can be $\nabla b_u = -e_{ui} + \beta_u b_u$, $\nabla b_i = -e_{ui} + \beta_v b_i$,

$$\nabla V_i. = -e_{ui} \left( \frac{1}{\sqrt{|\mathcal{P}_u \setminus \{i\}|}} \sum_{i' \in \mathcal{P}_u \setminus \{i\}} P_{i'} + \frac{1}{\sqrt{|\mathcal{E}_u^{(\ell)}|}} \sum_{j \in \mathcal{E}_u^{(\ell)}} E_j. \right) + \alpha_v V_i.$$

$$\nabla P_{i'} = -e_{ui} \frac{1}{\sqrt{|\mathcal{P}_u \setminus \{i\}|}} V_i. + \alpha_p P_{i'}, \quad i' \in \mathcal{P}_u \setminus \{i\}, \quad \text{and}$$

$$\nabla E_j. = -e_{ui} \frac{1}{\sqrt{|\mathcal{E}_u|}} V_i. + \alpha_e E_j., \quad j \in \mathcal{E}_u. \quad \text{Note that } e_{ui} = r_{ui} - \hat{r}_{ui} \text{ is the difference between the true preference and the estimated preference.}$$
Finally, we have the update rule for each corresponding $\theta \in \Theta^{(\ell)}$, 

$$\theta \leftarrow \theta - \gamma \nabla \theta,$$  

(7)

where $\gamma (\gamma > 0)$ is the learning rate.
Identification of Likely-to-Prefer Items

Once we have learned the model parameters $\Theta^{(\ell)}$, we can identify some likely-to-prefer items from $\mathcal{E}_u$ via the prediction rule in Eq.(4).

Specifically, for each examined item $j \in \mathcal{E}_u$ by user $u$, we estimate its preference score $\hat{r}_{uj}^{(\ell)}$, and then take $\tau|\mathcal{E}_u|$ examined items with the highest scores.

Note that $\tau$ ($0 < \tau \leq 1$) is a parameter for item selection, which is initialized as 1 in the beginning and is then gradually decreased via $\tau \leftarrow \tau \times 0.9$ so as to ignore some unlikely-to-prefer items.
Algorithm of TJSL (part 1)

- **Input**: Positive feedbacks $\mathcal{P}$, implicit examinations $\mathcal{E}$, iteration numbers $T, L, L_0$, and parameters $\rho, \alpha_v, \alpha_p, \alpha_e, \beta_u, \beta_v$.

- **Output**: Selected examinations $\mathcal{E}^{(\ell)}$ and learned models $\Theta^{(\ell)}$, $\ell = L - L_0 + 1, \ldots, L$.

The final prediction rule is $\hat{r}_{ui} = \sum_{\ell=L-L_0+1}^{L} \hat{r}_{ui}^{(\ell)}/L_0$, where $\hat{r}_{ui}^{(\ell)}$ is the prediction using the $\ell$th model parameters and data.
Algorithm of TJSL (part 2)

1: Let $\mathcal{E}(1) = \mathcal{E}$, $\tau = 1$
2: for $\ell = 1, \ldots, L$ do
3:   Initialize the model $\Theta^{(\ell)}$
4:       for $t = 1, \ldots, T$ do
5:           Randomly sample $\mathcal{A} \subset \mathcal{R} \setminus \mathcal{P}$ with $|\mathcal{A}| = \rho|\mathcal{P}|$
6:           for $t_2 = 1, \ldots, |\mathcal{P} \cup \mathcal{A}|$ do
7:               Randomly pick up $(u, i) \in \mathcal{P} \cup \mathcal{A}$
8:               Calculate $\hat{r}_{ui}^{(\ell)}$ via Eq.(4)
9:               Calculate $\nabla \theta, \theta \in \Theta^{(\ell)}$ via Eq.(6)
10:              Update $\theta, \theta \in \Theta^{(\ell)}$ via Eq.(7)
11:           end for
12:       end for
13:   if $\ell > L - L_0$ then
14:       Save the current model and data, i.e., $\Theta^{(\ell)}, \mathcal{E}^{(\ell)}$
15:   end if
16:   if $L > 1$ and $L > \ell$ then
17:       $\tau \leftarrow \tau \times 0.9$
18:       for $u \in \mathcal{U}$ do
19:           Select $\mathcal{E}_{u}^{(\ell+1)} \subseteq \mathcal{E}_{u}$ with $|\mathcal{E}_{u}^{(\ell+1)}| = \tau|\mathcal{E}_{u}|$
20:       end for
21:   end if
22: end for
when $L = L_0 = 1$, TJSL reduces to TJSL(1,1)
when $\mathcal{E} = \emptyset$, TJSL(1,1) reduces to $FISM_{rmse}$
when $T = 0$, $FISM_{rmse}$ reduces to PopRank
Datasets

**Table:** Description of the data sets used in the experiments, including numbers of users ($|U|$), items ($|I|$), positive feedbacks ($|P|$), implicit examinations ($|E|$), and test positive feedbacks ($|P^{te}|$).

| Data set      | $|U|$ | $|I|$ | $|P|$ | $|E|$ | $|P^{te}|$ |
|---------------|------|------|------|------|--------|
| MovieLens100K | 943  | 1682 | 9438 | 45285| 2153   |
| MovieLens1M   | 6040 | 3952 | 90848| 400083| 45075  |
| Alibaba2015   | 7475 | 5257 | 9290 | 62659| 2322   |

The data and code are available at

http://www.cse.ust.hk/~weikep/TL4HOCCF/
Baselines

- **BPR**: Bayesian Personalized Ranking
- **FISM$_{rmse}$**: Factored Item Similarity Model using RMSE loss
We use the statistics of positive feedbacks $\mathcal{P}$ to initialize the model parameters,

\[
\begin{align*}
    b_u &= \sum_{i=1}^{m} y_{ui} / m - \mu \\
    b_i &= \sum_{u=1}^{n} y_{ui} / n - \mu \\
    V_{ik} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
    P_{ik} &= (r - 0.5) \times 0.01, k = 1, \ldots, d \\
    E_{jk} &= (r - 0.5) \times 0.01, k = 1, \ldots, d
\end{align*}
\]

where $r$ ($0 \leq r < 1$) is a random variable, and $\mu = \sum_{u=1}^{n} \sum_{i=1}^{m} y_{ui} / n / m$. 
Parameter Configurations

We fix $\rho = 3$, $\gamma = 0.01$ and $d = 20$, and search the best values of the following parameters using $NDCG@5$,

- $\alpha_v = \alpha_p = \alpha_e = \beta_u = \beta_v \in \{0.001, 0.01, 0.1\}$
- $T \in \{100, 500, 1000\}$

For the iteration numbers, we fix $L = 10$ and $L_0 = 3$. 
### Observations

- \( \text{FISM}_{rmse} \) is close to or better than BPR, which shows the effectiveness of similarity learning and neighborhood-based prediction rule in \( \text{FISM}_{rmse} \) as compared with that of BPR; and

- TJSL further boosts the performance of \( \text{FISM}_{rmse} \) significantly via selection of examined items, which shows the usefulness of the selected examined items and the effectiveness of integrating them in a joint similarity learning manner.
We study a new and important recommendation problem called heterogeneous one-class collaborative filtering (HOCCF) containing positive feedbacks and implicit examinations.

We map the HOCCF problem to the transfer learning paradigm with target data (positive feedbacks) and auxiliary data (implicit examinations), and then design a novel transfer learning algorithm, i.e., transfer via joint similarity learning (TJSL).
Thank you!

- We thank the editors and reviewers for their expert comments and constructive suggestions.
- We thank the support of Natural Science Foundation of Guangdong Province No. 2014A030310268, National Natural Science Foundation of China No. 61502307, 61170077, Grant of Shenzhen City No. KQCX20140519103756206 and Natural Science Foundation of SZU No. 201436.