A Survey of Transfer Learning for Collaborative Recommendation with Auxiliary Data

Weike Pan
College of Computer Science and Software Engineering
Shenzhen University
panweike@szu.edu.cn

Part of this work was done when Weike Pan was a Ph.D. student under the supervision of Prof. Qiang Yang in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology.
Outline

• **Introduction**
  – Collaborative Recommendation
  – Auxiliary Data
  – A Transfer Learning View

• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
• Discussions and Future Directions
• Acknowledgement
Collaborative Recommendation

• A standard component in most Internet systems
  – E-commerce systems
  – Advertisement systems

• Limitation
  – Limited to users' feedbacks of explicit scores or implicit examinations

• Challenge
  – Sparsity, i.e., lack of users’ feedbacks
Introduction

Auxiliary Data

- Additional data: have the potential to help reduce the sparsity effect

<table>
<thead>
<tr>
<th>Content</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>user’s static profile of demographics, affiliations, etc.</td>
<td></td>
</tr>
<tr>
<td>item’s static description of price, brand, location, etc.</td>
<td></td>
</tr>
<tr>
<td>user-item pair’s user generated content (UGC), etc.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Context</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>user’s dynamic context of mood, health, etc.</td>
<td></td>
</tr>
<tr>
<td>item’s dynamic context of remaining quantities, etc.</td>
<td></td>
</tr>
<tr>
<td>user-item pair’s dynamic context of time, etc.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>user-user social network of friendship, etc.</td>
<td></td>
</tr>
<tr>
<td>item-item relevance network of taxonomy, etc.</td>
<td></td>
</tr>
<tr>
<td>user-item-user network of sharing items with friends, etc.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feedback</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>user’s feedback of rating on other items, etc.</td>
<td></td>
</tr>
<tr>
<td>item’s feedback of browsing by other users, etc.</td>
<td></td>
</tr>
<tr>
<td>user-item pair’s feedback of collection, etc.</td>
<td></td>
</tr>
</tbody>
</table>
A Transfer Learning View

- **Target data**: users' feedbacks
- **Auxiliary data**: other additional information

- **Focus**: how to achieve knowledge transfer from some auxiliary data to a target data, i.e., “how to transfer” in transfer learning [Pan and Yang, TKDE 2010]
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
  – Problem Definition
  – Categorization of Knowledge Transfer
  – A Generic Knowledge Transfer Framework
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
• Discussions and Future Directions
• Acknowledgement
Problem Definition

- **Target data**: a rating matrix $\mathbf{R} = [r_{ui}]^{n \times m}$
- **Auxiliary data**: $\mathbf{A}$, e.g., content, context, network and feedback
- **Goal**: predict the unobserved rating in $\mathbf{R}$ by transferring knowledge from $\mathbf{A}$

Transfer Learning for Collaborative Recommendation with Auxiliary Data (TL-CRAD)
Categorization of Knowledge Transfer

• Knowledge transfer algorithm *styles*
  – Adaptive knowledge transfer
  – Collective knowledge transfer
  – Integrative knowledge transfer

• Knowledge transfer *strategies*
  – Transfer via *prediction rule*
  – Transfer via *regularization*
  – Transfer via *constraint*
A Generic Knowledge Transfer Framework

• A generic framework for TL-CRAD

\[
\min_{\Theta, K} \mathcal{E}(\Theta, K|\mathbf{R}, \mathbf{A}) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t.} \Theta \in C(K, A)
\]

– \(\mathcal{E}(\Theta, K|\mathbf{R}, \mathbf{A})\): a loss function
– \(\mathcal{R}(\Theta|K, A), \mathcal{R}(K)\): two regularization terms
– \(\Theta \in C(K, A)\): a constraint

– \(\mathbf{R}\): target user-item rating matrix
– \(\mathbf{A}\): auxiliary data
– \(\mathbf{K}\): extracted knowledge from auxiliary data
– \(\Theta\): model parameters
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
  – Transfer via Regularization
  – Transfer via Constraint
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
• Discussions and Future Directions
• Acknowledgement
Adaptive Knowledge Transfer

Adaptive Knowledge Transfer

- **Adaptive knowledge transfer** aims to adapt the knowledge extracted from an auxiliary data domain to a target data domain. This is a directed knowledge transfer approach similar to traditional domain adaptation methods.
Transfer via Regularization

- **Instantiation** from the generic framework
  \[
  \min_{\Theta, K} \mathcal{E}(\Theta, K|\mathbf{R}, \mathbf{A}) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
  \]
  \[
  \rightarrow \quad \min_{\Theta} \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta|K)
  \]

- **Coordinate System Transfer (CST)** [Pan, Xiang, Liu and Yang, AAAI 2010]
  \[
  \mathcal{R}(\Theta|K) \rightarrow ||\mathbf{U} - \mathbf{U}'||_F^2 + ||\mathbf{V} - \mathbf{V}'||_F^2
  \]

- The two biased regularization terms are used to constrain the latent feature matrices of target data to be similar to that of auxiliary data.
- Biased regularization is a popular technique in domain adaptation.
Adaptive Knowledge Transfer

Transfer via Constraint

- **Instantiation** from the generic framework
  \[
  \min_{\Theta, K} \mathcal{E}(\Theta, K|R, A) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
  \]
  \[
  \quad \Rightarrow \quad \min_\Theta \mathcal{E}(\Theta|R), \text{ s.t. } \Theta \in \mathcal{C}(K)
  \]

**Example**

- **Codebook Transfer (CBT)** [Li, Yang and Xue, IJCAI 2009]
  \[
  \Theta \in \mathcal{C}(K)
  \]
  \[
  \quad \Rightarrow \quad B = \tilde{B}
  \]
  - The constraint on two codebooks is used to constrain cluster-level rating pattern of **target data** to be the same with that of **auxiliary data**.
  - Cluster-level rating pattern is a kind of **group behavior** with higher transferability.
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
  – Transfer via Constraint
• Integrative Knowledge Transfer
• Discussions and Future Directions
• Acknowledgement
Collective Knowledge Transfer

Collective Knowledge Transfer

- **Collective knowledge transfer** usually *jointly* learns the shared knowledge and unshared effect of the **target data** and the **auxiliary data** simultaneously. This is a **bi-directed** knowledge transfer approach with richer interactions similar to multi-task learning algorithms.

![Diagram showing collective knowledge transfer](image-url)
Transfer via Constraint

- **Instantiation** from the generic framework

\[
\min_{\Theta, K} \mathcal{E}(\Theta, K| R, A) + \mathcal{R}(\Theta| K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
\]

\[\Rightarrow \quad \min_{\Theta, K} \mathcal{E}(\Theta| R) + \mathcal{R}(\Theta) + \mathcal{E}(K| A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K)\]

**Example**

- **Collective Matrix Factorization (CMF)** [Singh and Gordon, KDD 2008]

\[
\Theta \in \mathcal{C}(K)
\]

\[\Rightarrow \quad V = \hat{V}\]

- The **constraint on two latent feature matrices** is used to enable knowledge transfer between the **target data** and the **auxiliary data**.

- The assumption that **same entities (e.g., users and/or items)** from the **target data** and the **auxiliary data** are associated with the **same latent factors** is quite universal.
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
  – Transfer via Prediction Rule
  – Transfer via Regularization
  – Transfer via Constraint
• Discussions and Future Directions
• Acknowledgement
Integrative knowledge transfer *incorporates* the raw *auxiliary data* as known knowledge into the learning task on the *target data*. It can be considered as an *embedded* knowledge transfer approach similar to feature engineering, information fusion and data integration methods.

![Diagram of Integrative Knowledge Transfer](image-url)
Transfer via Prediction Rule

- **Instantiation** from the generic framework

\[ \min_{\Theta,K} \mathcal{E}(\Theta, K|\mathbf{R}, \mathbf{A}) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t.} \Theta \in \mathcal{C}(K, A) \]

\[ \Rightarrow \min_{\Theta} \mathcal{E}(\Theta|\mathbf{R}, \mathbf{A}) + \mathcal{R}(\Theta) \]

**Example**

- Recommendation with Social Trust Ensemble (RSTE) [Ma, King and Lyu, TIST 2011]

\[ \hat{r}_{ui} = U_u \cdot V_i^T \]

\[ \Rightarrow \hat{r}_{ui} = \lambda U_u \cdot V_i^T + (1 - \lambda) \sum_{u' \in T_u^+} \tilde{e}_{u'i} U_{u'} \cdot V_i^T \]

- The **expanded prediction rule** is used to transfer social tastes from the **auxiliary data** to the **target data**.

- Integrating the auxiliary data via a revised prediction rule is a natural and effective knowledge transfer approach, though it may cause high time complexity.
Transfer via Regularization

- **Instantiation** from the generic framework

\[
\min_{\Theta, K} \mathcal{E}(\Theta, K|\mathbf{R}, \mathbf{A}) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
\]

\[\rightarrow \min_{\Theta} \mathcal{E}(\Theta|\mathbf{R}) + \mathcal{R}(\Theta|A)\]

**Example**

- Tag Informed Collaborative Filtering (TagiCoFi) [Zhen, Li and Yeung, RecSys 2009]

\[\mathcal{R}(\Theta|A) \rightarrow \sum_{u=1}^{n} \sum_{u' = 1}^{n} \hat{s}_{uu'} \| \mathbf{U}_u - \mathbf{U}_{u'} \|_F^2\]

- The manifold regularization term is used to constrain similar users in the **auxiliary data** to be similar in the latent space of the **target data**.
- Manifold regularization term and its variants are a popular technique in **semi-supervised machine learning**.
Transfer via Constraint

- **Instantiation** from the generic framework

\[
\min_{\Theta, K} \mathcal{E}(\Theta, K| R, A) + \mathcal{R}(\Theta| K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
\]

\[\Rightarrow \min_{\Theta} \mathcal{E}(\Theta| R) + \mathcal{R}(\Theta), \text{ s.t. } \Theta \in \mathcal{C}(A)\]

**Example**

- **Transfer by Integrative Factorization (TIF)** [Pan, Xiang and Yang, AAAI 2012]

\[
\Theta \in \mathcal{C}(A) \Rightarrow \hat{r}_{ui} \in \mathcal{C}(a_{ui}, b_{ui})
\]

- The **constraint** requires that the estimated preference by the learned model of the **target data** is **in the range of the corresponding uncertain rating** of the **auxiliary data**.

- Incorporating **auxiliary data** via **constraints** is quite flexible since auxiliary data can often be represented as some constraints.
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
• Discussions and Future Directions
  – Discussions
  – Future Directions
  – Conclusions
• Acknowledgement
Discussions and Future Directions

Discussions (1/2)

- **TL-CRAD**

![Diagram showing interaction between auxiliary and target data]

- The **interaction** between auxiliary data and target data becomes **richer** from adaptive, collective, to integrative algorithm styles, which are believed to enable **more effective knowledge transfer**.

- The **time complexity** may also **increase** from adaptive, collective, to integrative algorithm styles.
Discussions and Future Directions

Discussions (2/2)

• A generic framework

\[
\min_{\Theta, K} \mathcal{E}(\Theta, K|R, A) + \mathcal{R}(\Theta|K, A) + \mathcal{R}(K), \text{ s.t. } \Theta \in \mathcal{C}(K, A)
\]

• Summary

<table>
<thead>
<tr>
<th>Style</th>
<th>Strategy</th>
<th>Prediction rule</th>
<th>Regularization</th>
<th>Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(\min_{\Theta} \mathcal{E}(\Theta</td>
<td>R) + \mathcal{R}(\Theta</td>
<td>K))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>e.g., CBT [26], etc.</td>
</tr>
<tr>
<td>Adaptive</td>
<td></td>
<td>(\min_{\Theta, K} \mathcal{E}(\Theta</td>
<td>K, A) + \mathcal{R}(K))</td>
<td>(\min_{\Theta, K} \mathcal{E}(\Theta</td>
</tr>
<tr>
<td>Collective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrative</td>
<td></td>
<td>(\min_{\Theta} \mathcal{E}(\Theta</td>
<td>R, A) + \mathcal{R}(\Theta))</td>
<td>(\min_{\Theta} \mathcal{E}(\Theta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>e.g., FM [47], etc.</td>
<td>e.g., tagiCoFi [59], etc.</td>
<td>e.g., TIF [43], etc.</td>
</tr>
</tbody>
</table>

– Collective knowledge transfer via constraint and integrative knowledge transfer via prediction rule have recently received more attention.
Future Directions

• Heterogeneous Techniques Ensemble
  – Design some heterogeneous knowledge transfer algorithm styles and strategies in order to achieve a good balance among flexibility, effectiveness and efficiency.

• Heterogeneous Data Integration
  – Develop a unified framework for heterogeneous auxiliary data in a scalable and distributed way.

• Multi-Objective Recommendation
  – Design a sophisticated objective function with different evaluation metrics (e.g., accuracy, diversity, serendipity, quality of service) when exploiting the auxiliary data.

• Explanation and Security
  – Take auxiliary data as a source for explanation generation of the recommended items, and even for robustness against malicious attack or fake actions.
Conclusions

• Intelligent recommendation approaches (with one more category)

  – **Content-based Recommendation (CBR)**: promote an item based on the relevance between a candidate item and the active user's consumed items.

  – **Collaborative Recommendation (CR)**: recommend preferred items from similar-taste users.

  – **Transfer Learning for Collaborative Recommendation with Auxiliary Data (TL-CRAD)**: learn users’ preference via transferring knowledge from some auxiliary data (e.g., content, context, network and feedback) to a target data of users’ feedbacks.
Outline

• Introduction
• Collaborative Recommendation with Auxiliary Data
• Adaptive Knowledge Transfer
• Collective Knowledge Transfer
• Integrative Knowledge Transfer
• Discussions and Future Directions
• Acknowledgement
Thank You!

- Thanks for your attention!

- Thank Prof. Qiang Yang for advice and comments!

- Thank anonymous editors and reviewers for constructive suggestions!

- Thank the support of Natural Science Foundation of Guangdong Province, National Natural Science Foundation of China, and Natural Science Foundation of SZU!