Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation

Steffen Rendle\textsuperscript{1}, Lars Schmidt-Thieme

Information Systems and Machine Learning Lab
University of Hildesheim

Conference on Web Search and Data Mining (WSDM) 2010

\textsuperscript{1}currently on leave at Osaka University, Japan
Outline

1. Problem: Personalized Tag Recommendation

2. Optimization: Bayesian Personalized Ranking

3. Model: Pairwise Interaction Tensor Factorization

4. Evaluation

5. Conclusion
Personalized Tag Recommendation

Task: Recommend a user a (personalized) list of tags for a specific item.
Formalization

- $U$ ... users
- $I$ ... items
- $T$ ... tags
- $S \subseteq U \times I \times T$ ... observed tags
- $P_S = \{(u, i) | \exists t \in T : (u, i, t) \in S\}$ ... observed tagging posts
Formalization

All observations are positive!
For learning we also need negative examples.
Pairwise Training Data for Ranking

Advantages:

1. The pairwise preferences that should be recommended in the future, are treated as missing values.

2. Only within ‘observed’ posts, ranking constraints are derived.
Outline

1. Problem: Personalized Tag Recommendation

2. Optimization: Bayesian Personalized Ranking

3. Model: Pairwise Interaction Tensor Factorization

4. Evaluation

5. Conclusion
Optimization for Bayesian Personalized Ranking (BPR)

The Bayesian Personalized Ranking (BPR) optimization criterion for item recommendation [Rendle et al. UAI 2009] is:

$$\text{BPR-Opt}_{\text{item-rec}} := \ln p(\Theta > u)$$

$$= \sum_{u \in U} \sum_{i^+ \in I_u^+} \sum_{i^- \in I_u^-} \ln \sigma(\hat{y}_{u,i^+,i^-}) - \lambda \Theta ||\Theta||^2$$

Adapted to the task of tag recommendation, BPR-Opt is:

$$\text{BPR-Opt}_{\text{tag-rec}} := \ln p(\Theta > u,i)$$

$$= \sum_{(u,i) \in P} \sum_{t^+ \in T_{u,i}^+} \sum_{t^- \in T_{u,i}^-} \ln \sigma(\hat{y}_{u,i,t^+} - \hat{y}_{u,i,t^-}) - \lambda \Theta ||\Theta||^2$$

Next, we discuss how to model $\hat{y}_{u,i,t}$ with factorization models.
Outline

1. Problem: Personalized Tag Recommendation

2. Optimization: Bayesian Personalized Ranking

3. Model: Pairwise Interaction Tensor Factorization

4. Evaluation

5. Conclusion
Tucker Decomposition

\[ \hat{Y}_{u,i,t}^{\text{TD}} = \sum_{f_U=1}^{k_U} \sum_{f_i=1}^{k_I} \sum_{f_T=1}^{k_T} \hat{c}_{f_U,f_i,f_T} \cdot \hat{u}_{f_U} \cdot \hat{i}_{f_i} \cdot \hat{t}_{f_T} \]

- Complexity of model equation is cubic in \( k \).
- The tag recommender ‘RTF’ is based on this model [Rendle et al. KDD 2009].
Canonical Decomposition / Parallel Factor Analysis

\[ \hat{Y}_{u,i,t} = \sum_{f=1}^{k} \hat{u}_{u,f} \cdot \hat{i}_{i,f} \cdot \hat{t}_{t,f} \]

- Complexity of model equation is linear in \( k \).
- CD/ PARAFAC corresponds to TD with a static, diagonal core tensor.

[Harshman 1970, Carroll and Chang 1970]
Pairwise Interaction Tensor Factorization

\[ \hat{Y}_{u,i,t} = \sum_{f=1}^{k} \hat{u}_{u,f} \cdot \hat{t}^U_{t,f} + \sum_{f=1}^{k} \hat{i}_{i,f} \cdot \hat{t}^I_{t,f} = \langle \hat{u}_u, \hat{t}^U_t \rangle + \langle \hat{i}_i, \hat{t}^I_t \rangle \]

- Complexity of model equation is linear in \( k \).
- PITF models the two pairwise interactions \((U, T)\) and \((I, T)\) explicitly.
Expressiveness

\[ \mathcal{M}^{TD} \supset \mathcal{M}^{CD} \supset \mathcal{M}^{PITF} \]

- Tucker decomposition subsumes CD/PARAFAC which subsumes PITF.
- An advantage of CD/PARAFAC and PITF over TD is the linear complexity in \( k \).
- Theoretically, CD models should be as good or better as PITF. But under sparse settings, it might help to predefined the structure in advance.
Outline

1. Problem: Personalized Tag Recommendation
2. Optimization: Bayesian Personalized Ranking
3. Model: Pairwise Interaction Tensor Factorization
4. Evaluation
5. Conclusion
Evaluation

Datasets

<table>
<thead>
<tr>
<th>dataset</th>
<th>Users</th>
<th>Items</th>
<th>Tags</th>
<th>Triples</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>BibSonomy</td>
<td>116</td>
<td>361</td>
<td>412</td>
<td>10,148</td>
<td>2,522</td>
</tr>
<tr>
<td>Last.fm</td>
<td>2,917</td>
<td>1,853</td>
<td>2,045</td>
<td>219,702</td>
<td>75,565</td>
</tr>
<tr>
<td>ECML/PKDD DC09</td>
<td>1,185</td>
<td>22,389</td>
<td>13,276</td>
<td>248,494</td>
<td>63,628</td>
</tr>
</tbody>
</table>

Evaluated methods

- adapted PageRank for tag recommendation [Hotho et al. 2006]
- Folkrank [Hotho et al. 2006]
- HOSVD for tag recommendation [Symeonidis et al. 2008]
- RTF-TD (Ranking with Tensor Factorization) [Rendle 2009]
- upper bound for any non-personalized tag recommender
Recommendation quality

**BibSonomy**

- BPR–PITF 128
- BPR–CD 128
- RTF–TD 128
- FolkRank
- PageRank
- HOSVD

**Last.fm**

- BPR–PITF 64
- BPR–CD 64
- RTF–TD 64
- FolkRank
- PageRank
- HOSVD
- np\(_{\text{max}}\)
Learning Runtime

Last.fm: Prediction quality vs. learning runtime

- Top3 F-Measure vs. Learning runtime in days
- Top3 F-Measure vs. Learning runtime in minutes

- BPR–PITF 64
- BPR–CD 64
- RTF–TD 64
# ECML/PKDD Discovery Challenge 2009

<table>
<thead>
<tr>
<th>Rank</th>
<th>Method</th>
<th>Top-5 F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>BPR-PITF</strong> + adaptive list size</td>
<td>0.35594</td>
</tr>
<tr>
<td>-</td>
<td><strong>BPR-PITF</strong> <em>(not submitted)</em></td>
<td>0.345</td>
</tr>
<tr>
<td>2</td>
<td>Relational Classification [Marinho et al. 09]</td>
<td>0.33185</td>
</tr>
<tr>
<td>3</td>
<td>Content-based [Lipczak et al. 09]</td>
<td>0.32461</td>
</tr>
<tr>
<td>4</td>
<td>Content-based [Zhang et al. 09]</td>
<td>0.32230</td>
</tr>
<tr>
<td>5</td>
<td>Content-based [Ju and Hwang 09]</td>
<td>0.32134</td>
</tr>
<tr>
<td>6</td>
<td>Personomy translation [Wetzker et al. 09]</td>
<td>0.32124</td>
</tr>
</tbody>
</table>

...  ...  

Task 2: ECML/ PKDD Challenge 2009,
http://www.kde.cs.uni-kassel.de/ws/dc09/results
Outline

1. Problem: Personalized Tag Recommendation

2. Optimization: Bayesian Personalized Ranking

3. Model: Pairwise Interaction Tensor Factorization

4. Evaluation

5. Conclusion
Conclusion

- PITF explicitly models the two pairwise interactions among user/tags and item/tags.

- Even though CD/PARAFAC and TD subsume PITF, this does not mean that they are guaranteed to generate better recommendations under sparsity!

- Empirically PITF outperformed all approaches on the Last.fm dataset and the ECML/PKDD Discovery Challenge 2009.