Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation

Steffen Rendle[1], Lars Schmidt-Thieme

Information Systems and Machine Learning Lab
University of Hildesheim

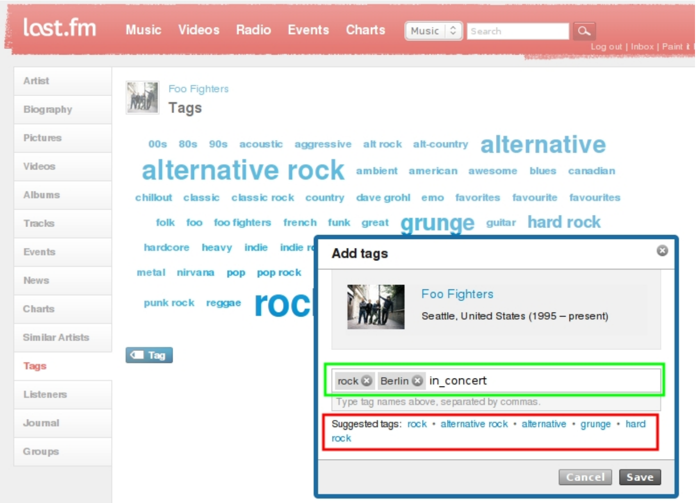
Conference on Web Search and Data Mining (WSDM) 2010

[1]currently on leave at Osaka University, Japan

Outline

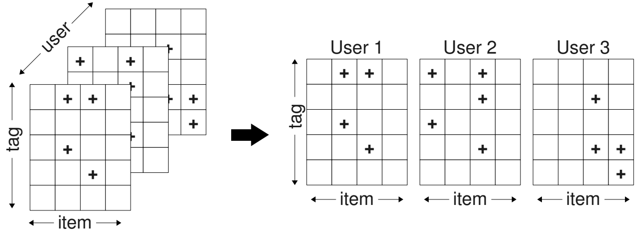
1. Problem: Personalized Tag Recommendation
2. Optimization: Bayesian Personalized Ranking
3. Model: Pairwise Interaction Tensor Factorization
4. Evaluation
5. Conclusion

Personalized Tag Recommendation



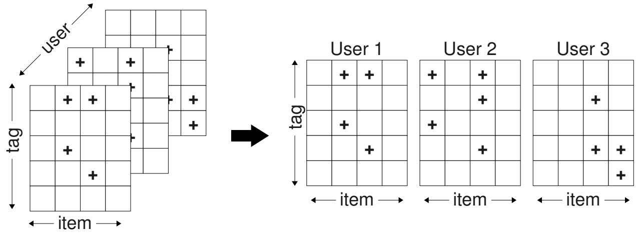
Task: Recommend a user a (personalized) list of tags for a specific item.

Formalization



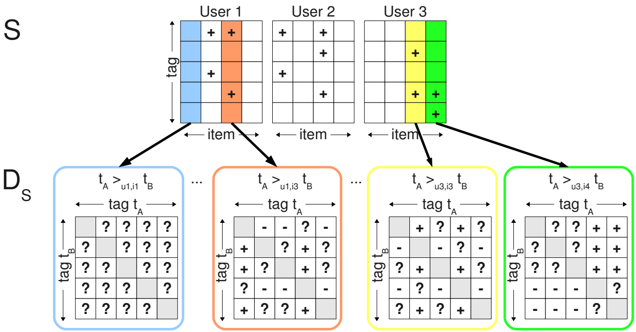
- U ... users
- I ... items
- T ... tags
- $S \subseteq U \times I \times T$ ... observed tags
- $P_S = \{(u, i) | \exists t \in T : (u, i, t) \in S\}$ ... observed tagging posts

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:

$$\begin{aligned}\text{BPR-OPT}_{\text{item-rec}} &:= \ln\, p(\Theta|>_u) \\ &= \sum_{u \in U} \sum_{i^+ \in I_u^+} \sum_{i^- \in I_u^-} \ln \sigma(\hat{y}_{u,i^+,i^-}) - \lambda_\Theta ||\Theta||^2\end{aligned}$$

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

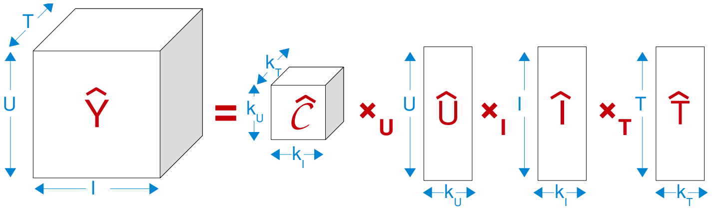
$$\begin{aligned}\text{BPR-OPT}_{\text{tag-rec}} &:= \ln\, p(\Theta|>_{u,i}) \\ &= \sum_{(u,i) \in P_S} \sum_{t^+ \in T_{u,i}^+} \sum_{t^- \in T_{u,i}^-} \ln \sigma(\hat{y}_{u,i,t^+} - \hat{y}_{u,i,t^-}) - \lambda_\Theta ||\Theta||^2\end{aligned}$$

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

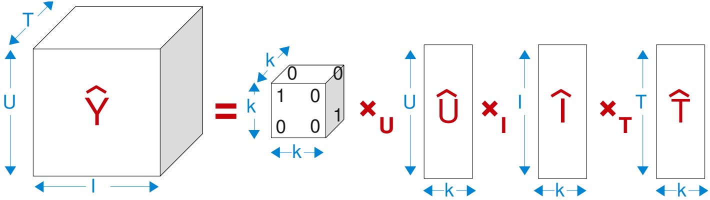


$$\hat{y}^{\mathrm{TD}}_{u,i,t} = \sum_{f_U=1}^{k_U} \sum_{f_I=1}^{k_I} \sum_{f_T=1}^{k_T} \hat{c}_{f_U,f_I,f_T} \cdot \hat{u}_{u,f_U} \cdot \hat{i}_{i,f_I} \cdot \hat{t}_{t,f_T}$$

- Complexity of model equation is cubic in k.
- The tag recommender 'RTF' is based on this model [Rendle et al. KDD 2009].

[Tucker 1966]

Canonical Decomposition / Parallel Factor Analysis

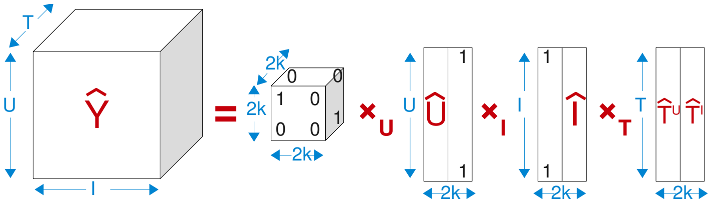


$$\hat{y}^{\text{CD}}_{u,i,t} = \sum_{f=1}^{k} \hat{u}_{u,f} \cdot \hat{i}_{i,f} \cdot \hat{t}_{t,f}$$

- 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



$$\hat{y}^{\text{PITF}}_{u,i,t} = \sum_{f=1}^{k} \hat{u}_{u,f} \cdot \hat{t}^{U}_{t,f} + \sum_{f=1}^{k} \hat{i}_{i,f} \cdot \hat{t}^{I}_{t,f} = \langle \hat{\mathbf{u}}_u, \hat{\mathbf{t}}^{U}_t \rangle + \langle \hat{\mathbf{i}}_i, \hat{\mathbf{t}}^{I}_t \rangle$$

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

Expressiveness

$$\mathcal{M}^{\text{TD}} \supset \mathcal{M}^{\text{CD}} \supset \mathcal{M}^{\text{PITF}}$$

- 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 predefine 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

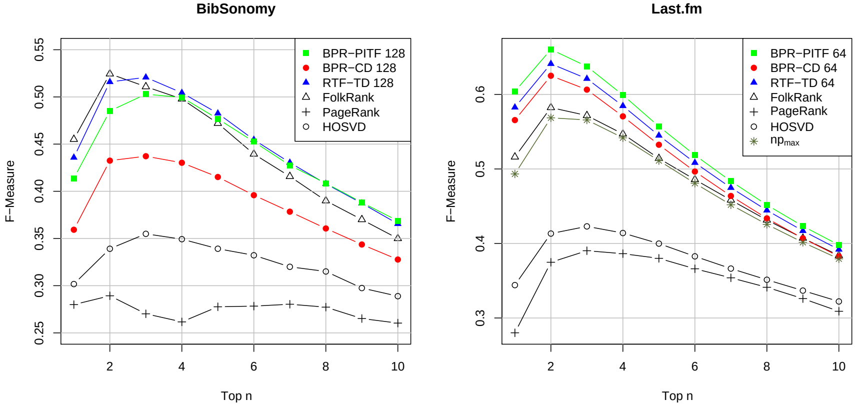
Datasets

dataset	Users $\|U\|$	Items $\|I\|$	Tags $\|T\|$	Triples $\|S\|$	Posts $\|P_S\|$
BibSonomy	116	361	412	10,148	2,522
Last.fm	2,917	1,853	2,045	219,702	75,565
ECML/PKDD DC09	1,185	22,389	13,276	248,494	63,628

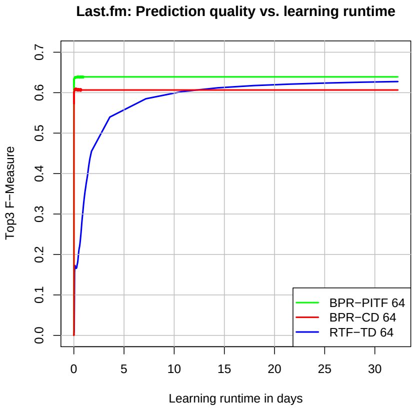
Evaluated methods

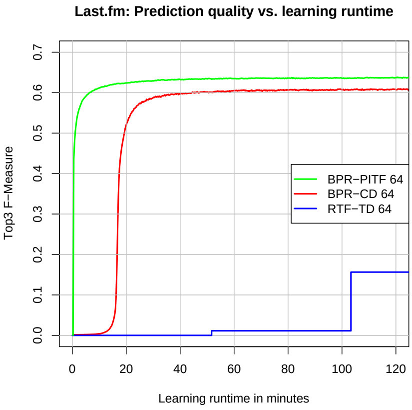
- adapted PageRank for tag recommendation [Hotho et al. 2006]
- Folkrank [Hotho et al. 2006]
- HOSVD for tag recommendation [Symeonidis et al. 2008]
- RTF-TD (Ranking with Tensor Factorization) [Rendle 2009]
- upper bound for any non-personalized tag recommender

Recommendation quality



Learning Runtime





ECML/PKDD Discovery Challenge 2009

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

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.