Factorizing Personalized Markov Chains for Next-Basket Recommendation

Steffen Rendle, Christoph Freudenthaler, Lars Schmidt-Thieme

Information Systems and Machine Learning Lab
University of Hildesheim

World Wide Web Conference (WWW)
April 29, 2010

Recommender Systems

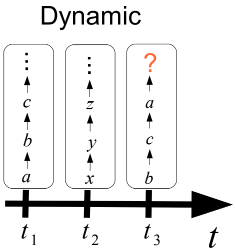
Target: Prediction of yet unknown items to interested users

- **Example 1:** Best films (e.g. Netflix)
- **Example 2:** Most appropriate tag (e.g. photos, videos)

Problem: Find an optimally sorted list of items (= ranking!)

- **Sorting criterion 1:** Ratings
- **Sorting criterion 2:** Binary selection probabilities (purchases, views,. . .)

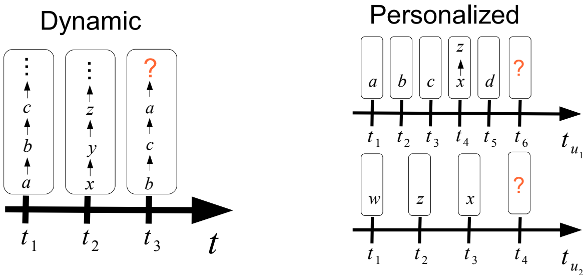
Recommender Systems



Existing approaches:

- **Dynamical Models:** Context matters (e.g. songs)

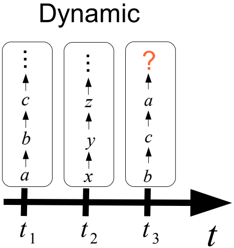
Recommender Systems

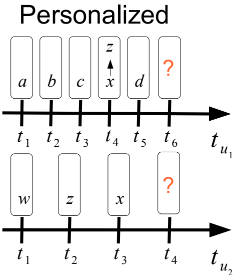


Existing approaches:

- **Dynamical Models:** Context matters (e.g. songs)
- **Personalization Models:** User profile matters (e.g. films)

Recommender Systems





Existing approaches:

- **Dynamical Models:** Context matters (e.g. songs)
- **Personalization Models:** User profile matters (e.g. films)

Note: Standard problems are almost static (Netflix: movie)

Recommender Systems

Dynamic	Personalized	Model	References
yes	no	Frequent Sequence Mining	Zimdars et al, UAI 2001
yes	no	Markov Chain Models (MC)	Shani et al, JMLR 2005
no	yes	Matrix Factorization (MF)	Koren, KDD 2008
...	...	...	...
yes	yes	Factor. Personal. MC (FPMC)	WWW2010

Our approach: Factorizing Personalized Markov Chains (FPMC)

- Generalization of existing approaches (MF and MC $\rightarrow$ FPMC)
- Combination of **personalization** and **sequential information** with Factorization Models
- Factorization models alleviate sparsity problem of ML-estimators (for probability density estimation)
- Item prediction = ranking $\rightarrow$ change in loss function

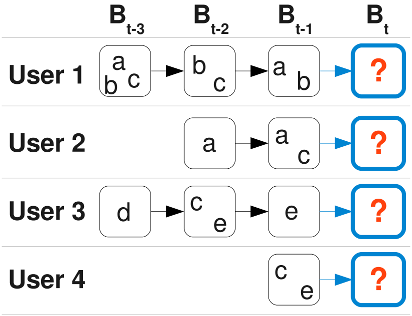
Outline

1. Dynamic Problem: Sequential Basket Data
2. Model: Factorizing Personalized Markov Chains (FPMC)
   Personalized Markov Chains for Sets
   Factorizing Transition Graphs
3. FPMC for Item Recommendation
4. Evaluation

Outline

1. Dynamic Problem: Sequential Basket Data
2. Model: Factorizing Personalized Markov Chains (FPMC)
   Personalized Markov Chains for Sets
   Factorizing Transition Graphs
3. FPMC for Item Recommendation
4. Evaluation

Next Basket Recommendation



Task: Which basket should be recommended to the user?

Problem properties:

- Special type of item recommendation (baskets ↔ sets of items)
- Reoccurring items
- Possible scenario: online store

Formalization

- $U = \{u_1, \ldots, u_{|U|}\}$ ... users
- $I = \{i_1, \ldots, i_{|I|}\}$ ... items
- $B \subset I$ ... basket
- $\mathcal{B}^u = (B_1^u, \ldots, B_{t_u-1}^u)$ ... historical baskets of user u
- $\mathcal{B} = \{\mathcal{B}^{u_1}, \ldots, \mathcal{B}^{u_{|U|}}\}$ ... all historical data

Outline

1. Dynamic Problem: Sequential Basket Data

2. Model: Factorizing Personalized Markov Chains (FPMC)
   Personalized Markov Chains for Sets
   Factorizing Transition Graphs

3. FPMC for Item Recommendation

4. Evaluation

Markov Chains for Baskets

Remember:

- Markov chain of order m

$$p(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_{t-m} = x_{t-m})$$

where $X_t, \dots, X_{t-m}$ are random variables and $x_t, \dots, x_{t-m}$ their realizations.

Markov Chains for Baskets

Remember:

- Markov chain of order m

$$p(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_{t-m} = x_{t-m})$$

  where $X_t, \ldots, X_{t-m}$ are random variables and $x_t, \ldots, x_{t-m}$ their realizations.

- . . . has $|X|^{m+1}$ parameters!

Markov Chains for Baskets

Markov chain for basket data of order 1:

$$p(B_t | B_{t-1})$$

Problem:

- Variables (baskets) are defined over $\mathcal{P}(I)$.
- Size of the state space is $2^{|I|}$.

Markov Chains for Baskets

Transition probabilities over binary purchasing events:

$$p(i \in B_t | \ell \in B_{t-1}) =: A_{\ell,i}$$

Predictions using simplified Markov Chains:

$$p(i \in B_t | B_{t-1}) := \frac{1}{|B_{t-1}|} \sum_{\ell \in B_{t-1}} p(i \in B_t | \ell \in B_{t-1})$$

- $|I|$ binary variables.
- Matrix A has size $|I|^2$
- A is no stochastic transition matrix anymore

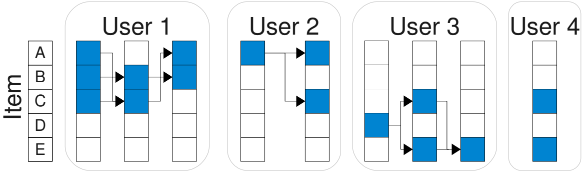
Markov Chains for Baskets

Maximum likelihood estimator for $A_{\ell,i}$ given $\mathcal{B}$ is:

$$\hat{A}_{\ell,i} = \hat{p}(i \in B_t | \ell \in B_{t-1}) = \frac{\hat{p}(i \in B_t \wedge \ell \in B_{t-1})}{\hat{p}(\ell \in B_{t-1})} =$$

$$= \frac{|\{(B_t, B_{t-1}) : i \in B_t \wedge \ell \in B_{t-1}\}|}{|\{(B_t, B_{t-1}) : \ell \in B_{t-1}\}|}$$

Example:



Transition Matrix

from \ to	A	B	C	D	E	#
A	0.5	0.5	1	0	0	2
B	0.5	1	0.5	0	0	2
C	0.3	0.7	0.3	0	0.3	3
D	0	0	1	0	1	1
E	0	0	0	0	1	1

Personalized Markov Chains

Each user has an own Markov Chain, i.e. the transition probabilities are individual:

$$A_{u,\ell,i} := p(i \in B_t^u | \ell \in B_{t-1}^u)$$

And:

$$p(i \in B_t^u | B_{t-1}^u) := \frac{1}{|B_{t-1}^u|} \sum_{\ell \in B_{t-1}^u} p(i \in B_t^u | \ell \in B_{t-1}^u)$$

$A \in [0,1]^{|U| \times |I| \times |I|}$ is a transition tensor.

Problems of personalized ML-Estimation:

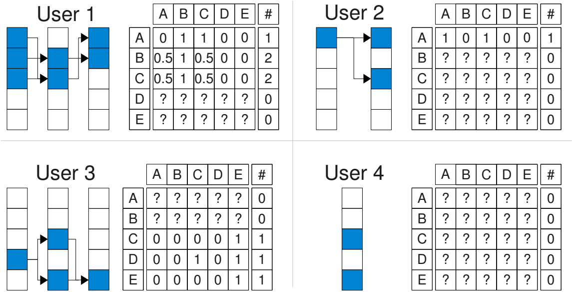
- Data sparsity explodes
- Theoretical properties only valid for large data sets

Personalized Markov Chains

Maximum likelihood estimator for $A_{u,\ell,i}$ given $\mathcal{B}$ is:

$$\hat{A}_{u,\ell,i} = \hat{p}(i \in B_t^u | \ell \in B_{t-1}^u) = \frac{|\{(B_t^u, B_{t-1}^u) : i \in B_t^u \wedge \ell \in B_{t-1}^u\}|}{|\{(B_t^u, B_{t-1}^u) : l \in B_{t-1}^u\}|}$$

Example:



User 1

	A	B	C	D	E	#
A	0	1	1	0	0	1
B	0.5	1	0.5	0	0	2
C	0.5	1	0.5	0	0	2
D	?	?	?	?	?	0
E	?	?	?	?	?	0

User 2

	A	B	C	D	E	#
A	1	0	1	0	0	1
B	?	?	?	?	?	0
C	?	?	?	?	?	0
D	?	?	?	?	?	0
E	?	?	?	?	?	0

User 3

	A	B	C	D	E	#
A	?	?	?	?	?	0
B	?	?	?	?	?	0
C	0	0	0	0	1	1
D	0	0	1	0	1	1
E	0	0	0	0	1	1

User 4

	A	B	C	D	E	#
A	?	?	?	?	?	0
B	?	?	?	?	?	0
C	?	?	?	?	?	0
D	?	?	?	?	?	0
E	?	?	?	?	?	0

Factorization Model

$\hat{A}$ is modeled by the pairwise interaction tensor factorization:

$$\hat{A}_{u,\ell,i} := \langle v_u^{U,I}, v_i^{I,U} \rangle + \langle v_i^{I,L}, v_\ell^{L,I} \rangle + \langle v_u^{U,L}, v_\ell^{L,U} \rangle$$

or equivalently:

$$\hat{A}_{u,\ell,i} := \sum_{f=1}^{k_{U,I}} v_{u,f}^{U,I}\, v_{i,f}^{I,U} + \sum_{f=1}^{k_{I,L}} v_{i,f}^{I,L}\, v_{\ell,f}^{L,I} + \sum_{f=1}^{k_{U,L}} v_{u,f}^{U,L}\, v_{\ell,f}^{L,U}$$

Note: also other models like Tucker Decomposition or PARAFAC would be possible.

Factorizing Personalized Markov Chains (FPMC)

In total the FPMC model is:

$$\hat{p}(i \in B_t^u | B_{t-1}^u) = \langle v_u^{U,I}, v_i^{I,U} \rangle + \frac{1}{|B_{t-1}^u|} \sum_{\ell \in B_{t-1}^u} \left(\langle v_i^{I,L}, v_\ell^{L,I} \rangle + \langle v_u^{U,L}, v_\ell^{L,U} \rangle \right)$$

As A is modelled by a low rank approximation $\hat{A}$:

- The parameters $A_{u,\ell,i}$ are not independent of each other.
- Parameters are estimated from many data points.
- Generalization to non observed elements (regularization).

Outline

1. Dynamic Problem: Sequential Basket Data

2. Model: Factorizing Personalized Markov Chains (FPMC)
   Personalized Markov Chains for Sets
   Factorizing Transition Graphs

3. FPMC for Item Recommendation

4. Evaluation

Item Recommendation

The task is to estimate a personalized, dynamic ranking over the items:

$$<_{u,t} \subset I^2$$

The maximum a posteriori estimator is:

$$\operatorname*{argmax}_{\Theta} \sum_{u \in U} \sum_{B_t \in \mathcal{B}^u} \sum_{i \in B_t} \sum_{j \notin B_t} \ln \sigma(\hat{x}_{u,t,i} - \hat{x}_{u,t,j}) - \lambda_{\Theta} ||\Theta||_F^2$$

where $\hat{x}_{u,t,i} = \hat{p}(i \in B_t^u | B_{t-1}^u)$ and Θ are the factorization matrices.

Learning algorithm: Stochastic gradient descent

Expressiveness

For item recommendation with FPMC, the (U, L) interaction can be dropped. And FPMC can be rewritten as:

$$\hat{x}^{\text{FPMC}}_{u,t,i} = \hat{x}^{\text{MF}}_{u,t,i} + \hat{x}^{\text{FMC}}_{u,t,i}$$

where $\hat{x}^{\text{MF}}_{u,t,i}$ is the standard matrix factorization model:

$$\hat{x}^{\text{MF}}_{u,t,i} = \langle v^{U,I}_u, v^{I,U}_i \rangle$$

and $\hat{x}^{\text{FMC}}_{u,t,i}$ is the non-personalized factorized Markov chain:

$$\hat{x}^{\text{FMC}}_{u,t,i} := \frac{1}{|B_{t-1}|} \sum_{\ell \in B_{t-1}} \langle v^{I,L}_i, v^{L,I}_\ell \rangle$$

Thus FPMC subsumes both the MF and MC models.

Expressiveness

Note:

$$\hat{x}^{\text{FPMC}}_{u,t,i} = \hat{x}^{\text{MF}}_{u,t,i} + \hat{x}^{\text{FMC}}_{u,t,i}$$

- is more than simple ensembling (joint learning)
- comes from the chosen factorization model and ranking criterion

Outline

1. Dynamic Problem: Sequential Basket Data
2. Model: Factorizing Personalized Markov Chains (FPMC)
   Personalized Markov Chains for Sets
   Factorizing Transition Graphs
3. FPMC for Item Recommendation
4. Evaluation

Dataset

Evaluation is done on a sparse and dense subset of a Drug store:

dataset	users $|U|$	items $|I|$	baskets	$\overline{|B|}$	avg. baskets per user
sparse	71,602	7,180	233,476	11.3	3.2
dense	10,000	1,002	90,655	9.2	9.0

Properties of the MC transition matrix estimated by the MLE:

datase	missing values	non-zero	zero
sparse	1,041,100 (2.0%)	6,234,371 (12.1 %)	44,276,929 (85.9%)
dense	0 (0.0%)	889,419 (88.6 %)	114,585 (11.4%)

- **Sparse data set:** 10-core on users and items
- **Dense data set:** most bought items, most active users

Evaluation results

Measures:

- F-measure, AUC, HLU

Evaluation results

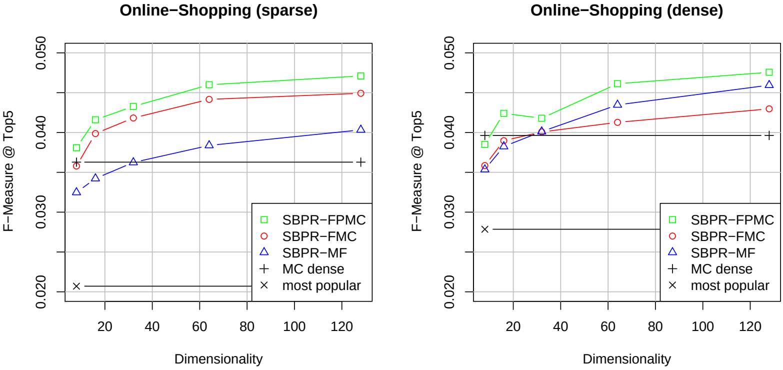
Measures:

- F-measure, AUC, HLU

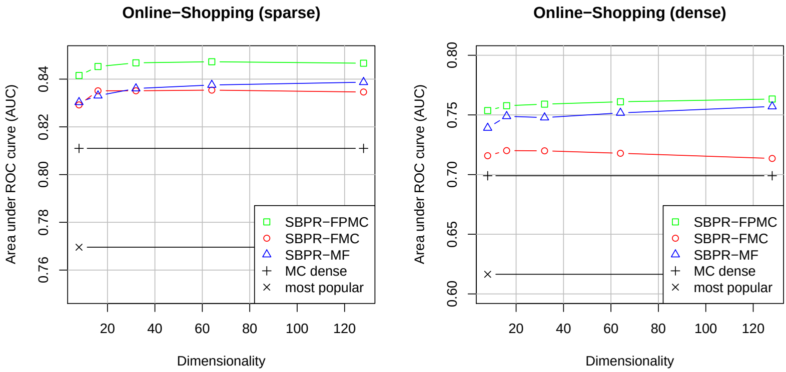
Compared Models:

	personalized	non-personalized
dynamic	FPMC	FMC, MC
static	MF	most popular

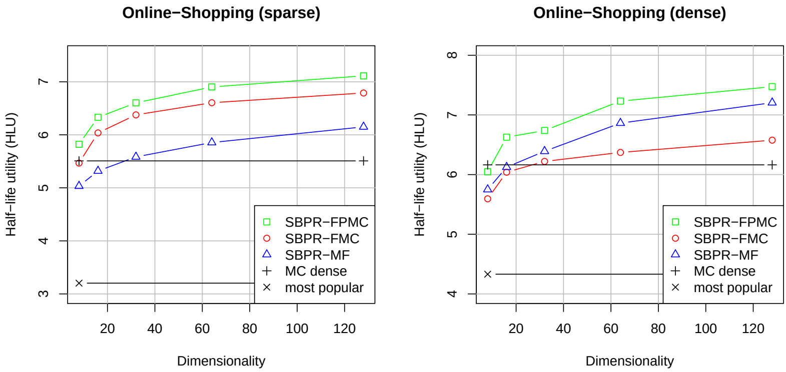
Prediction Quality



Prediction Quality



Prediction Quality



Conclusion

- FPMC combines the advantages of personalization, sequence information and factorization models
- Personalized Markov chains have individual transition graphs per user.
- Sparseness is solved by factorizing the transition cube (outperform Maximum Likelihood Models)
- FPMC subsumes Matrix Factorization and (Fact.) Markov chain model
- FPMC outperforms all evaluated models