Next Basket Recommendation

Boyoung Kim

Seoul National University

July 3, 2017

Introduction : Next Basket Recommendation

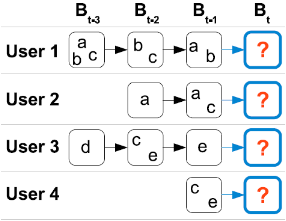


Figure 1: Sequential basket data with four users and five items

Which basket should be recommended to the user?

Properties :

- Special type of item recommendation (sets of items)
- Reoccuring items
- Possible scenario : online store

Factorizing Personalized Markov Chains for Next-Basket Recommendation

International World Wide Web Conference 2010

Steffen Rendel, Christoph Freudenthalen, Lars Schmidt-Thieme

Problem Formulation

- $U = \{u_1, \ldots, u_{|U|}\}$: users
- $I = \{i_1, \ldots, i_{|I|}\}$: items
- $\mathcal{B}^u = (B_1^u, \ldots, B_{t_u-1}^u)$: historical baskets of user u with $B_t^u \subseteq I$
- $\mathcal{B} = \{\mathcal{B}^{u_1}, \ldots, \mathcal{B}^{u_{|U|}}\}$: purchase history of all users

Markov Chains for Baskets

- A sequence of random variables $X_1, X_2, \ldots$ is a Markov chain of order m if

$$p(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_1 = x_1)$$
$$= p(X_t = x_t | X_{t-1} = x_{t-1}, \ldots, X_{t-m} = x_{t-m}).$$

- A personalized Markov chain of order $m = 1$ for basket problem is :

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

Markov Chains for Baskets

- We model transitions over $|I|$ binary variables that describe a basket:

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

- Predictions using simplified Markov chains :

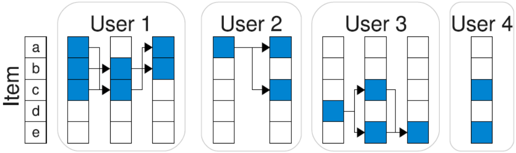
$$p(i \in B_t | B_{t-1}) := \frac{1}{|B_{t-1}|} \sum_{l \in B_{t-1}} p(i \in B_t | l \in B_{t-1})$$

Markov Chains for Baskets

- Maximum likelihood estimator for $a_{l,i}$ given $\mathcal{B}$ is :

$$
\begin{aligned}
\hat{a}_{l,i} &= \hat{p}(i \in B_t | l \in B_{t-1}) = \frac{\hat{p}(i \in B_t \wedge l \in B_{t-1})}{\hat{p}(l \in B_t)} \\
&= \frac{|(B_t, B_{t-1}) : i \in B_t \wedge l \in B_{t-1}|}{|(B_t, B_{t-1}) : l \in B_{t-1}|}
\end{aligned}
$$

- 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

Figure 2: Non-personalized Markov chain

Personalized Markov Chains

- We represent each MC by the transitions over items, but now user-specific:

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

- The prediction depends only on the users transitions:

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

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

ν Problems of personalized ML-Estimation:

- Many of the parameters cannot be estimated.
- Theoretical properties are easily fail since the data is extremely sparse.

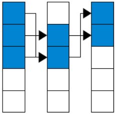
Personalized Markov Chains

- Maximum likelihood estimator for $a_{u,l,i}$ given $\mathcal{B}$ is :

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

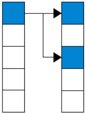
- Example :

User 1



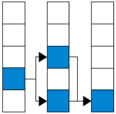
	a	b	c	d	e	#
a	0	1	1	0	0	1
b	0.5	1	0.5	0	0	2
c	0.5	1	0.5	0	0	2
d	?	?	?	?	?	0
e	?	?	?	?	?	0

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 of the Transition Cube

- $\hat{\mathcal{A}}$ is modeled by the pairwise interactions :

$$\hat{a}_{u,l,i} := \langle v_u^{U,I}, v_i^{I,U} \rangle + \langle v_i^{I,L}, v_l^{L,I} \rangle + \langle v_u^{U,L}, v_l^{L,U} \rangle$$

- or equivalently :

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

where $V^{U,I} \in \mathbb{R}^{|U| \times k_{U,I}}$, $V^{I,U} \in \mathbb{R}^{|I| \times k_{U,I}}$, $V^{I,L} \in \mathbb{R}^{|I| \times k_{I,L}}$, $V^{L,I} \in \mathbb{R}^{|I| \times k_{I,L}}$, $V^{U,L} \in \mathbb{R}^{|U| \times k_{U,L}}$ and $V^{L,U} \in \mathbb{R}^{|I| \times k_{U,L}}$.

Factorizing Personalized Markov Chains (FPMC)

- In total the FPMC model is:

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

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

  - Parameters are estimated from many data points.

Item Recommendation with FPMC

- Bayesian Personalized Ranking (BPR) [1]
- Assumption : a user prefers an item in basket at a specific time than any other items apart frome those in the basket.
- For any item i in basket and any item j not in basket, we need to maximize

$$p(i >_{u,t} j|\Theta) = p(\hat{x}_{u,t,i} >_{\mathbb{R}} \hat{x}_{u,t,j}|\Theta)$$

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

- The objective function can be written as follows :

$$J = -\sum_{u \in U} \sum_{B_t \in \mathcal{B}^u} \sum_{i \in B_t} \sum_{j \notin B_t} ln\ \sigma(\hat{x}_{u,t,i} - \hat{x}_{u,t,j}) + \frac{\lambda_\Theta}{2} \parallel \Theta \parallel_{\mathcal{F}}^2$$

[1] S. Rendle, et al. *Bpr : Bayesian personalized ranking from implicit feedback*. In UAI, 2009.

Expressiveness

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

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

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

$$\hat{x}_{u,t,i}^{MF} = \langle v_u^{U,I}, v_i^{I,U} \rangle$$

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

$$\hat{x}_{u,t,i}^{FMC} = \frac{1}{|B_{t-1}|} \sum_{l \in B_{t-1}} \langle v_i^{I,L}, v_l^{L,I} \rangle.$$

- Thus FPMC subsumes both the MF and MC models.

Experimental results

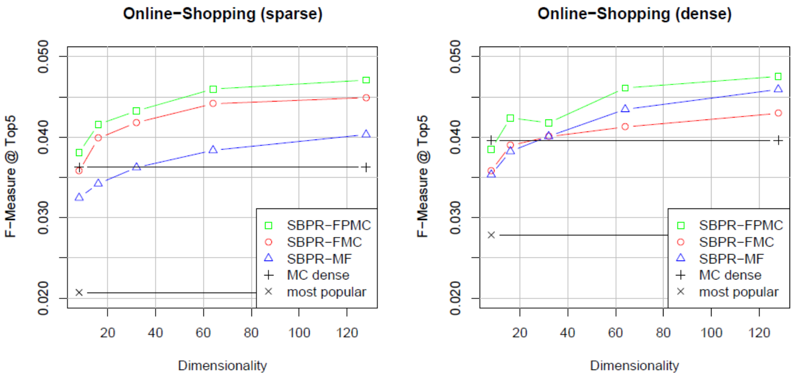


Figure 4: Comparison of FPMC to a factorized Markov chain (FMC), matrix factorization (MF) a standard dense Markov chain (MC dense) learned with Maximum Likelihood and the baseline most-popular.

A Dynamic Recurrent Model for Next Basket Recommendation

International World Wide Web Conference 2010

Feng Yu, et al.

Intuition of DREAM

- MC based methods can only model local sequential behaviors between two adjacent baskets.

  ⇒ Deep RNN architecture can obtain some global sequential features for all users from historical transaction data.

Problem Formulation

- $N \in \mathbb{R}^{I \times k}$: represantations of items.
- $B^u = \{B^u_{t_1}, B^u_{t_2}, \ldots\}$
  where $B^u_{t_i}$ is a basket of items purchased by user u at time t_i.
- $B^u_{t_i} = \{n^u_{t_{i,j}} \in \mathbb{R}^k \mid j = 1, \ldots, |B^u_{t_i}|\}$
- $n^u_{t_{i,j}}$: latent representation of the j-th item in B_{t_i}

A Dynamic Recurrent Model (DREAM)

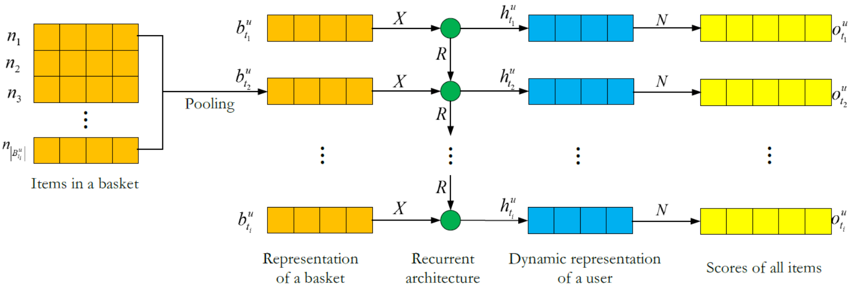


Figure 5: The framework of DREAM.

A Dynamic Recurrent Model (DREAM)

- *Max pooling* operation

  $b^u_{t_i,k}$ represents latent basket representation of the user's basket at time t_i :

$$b^u_{t_i,k} = max(n^u_{t_{i,1,k}}, n^u_{t_{i,2,k}}, \ldots, n^u_{t_{i,|B^u_{t_i}|,k}})$$

  where $n^u_{t_{i,j,k}}$ is the k-th value of $n^u_{t_{i,j}}$.

A Dynamic Recurrent Model (DREAM)

- Hidden features

  $h^u_{t_i}$ is dynamic representation of user u at time t_i :

$$h^u_{t_i} = f(Xb^u_{t_i} + Rh^u_{t_i-1})$$

  where $R \in \mathbb{R}^{k \times k}$ propagates sequential signals between every two adjacent hidden states, $X \in \mathbb{R}^{k \times k}$ is a transition matrix between latent vector representation of baskets and user's interest and f is an activation function (sigmoid).

- Output

  $o^u_{t_i} \in \mathbb{R}^I$ is scores of all items for user u :

$$o^u_{t_i} = Nh^u_{t_i}$$

  where $N \in \mathbb{R}^{I \times k}$ is item matrix.

Experimental results

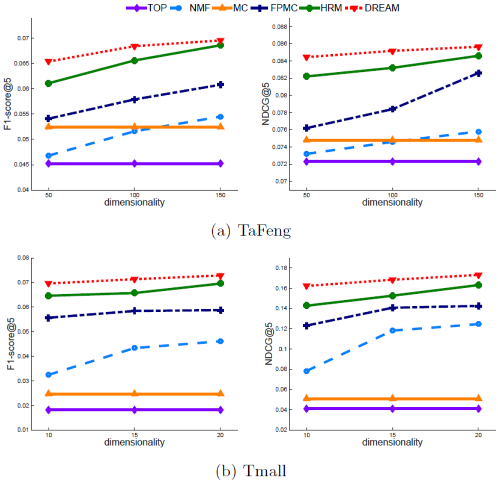


Figure 6: Experiment Results of different methods on two datasets.