BPR: Bayesian Personalized Ranking from Implicit Feedback

Steffen Rendle, Christoph Freudenthaler, Zeno Gantner and Lars Schmidt-Thieme

Basic Ideas(1)

■ Recommendation system

Users가 어떤 items에게 점수를 주었을 때, Users가 아직 점수를 주지않은 items들의 점수를 어떻게 예측하는가

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

[Figure 1] A matrix of user/item ratings

Basic Ideas(2)

■ Explicit Feedback
- Item Rating
- Item Ranking

U/I	Item 1	Item 2	Item 3	...	Item I
User 1		5			3
User 2	2		1		
...		3		4	
User U	1		4		5

[Figure 2.1] User preference to rated items (Explicit Feedback)

■ Implicit Feedback
- Bought (or not)
- Click (or not)

U/I	Item 1	Item 2	Item 3	...	Item I
User 1		O			O
User 2	O		O		
...		O		O	
User U	O		O		O

[Figure 2.2] User preference with binary labels (Implicit Feedback)

Basic Ideas(2)

■ ~~Explicit Feedback~~

- Item Rating
- Item Ranking

■ Implicit Feedback

- Bought (or not)
- Click (or not)

Very Sparse Dataset

U/I	Item 1	Item 2	Item 3	...	Item I
User 1		5			3
User 2	2		1		
...		3		4	
User U	1		4		5

[Figure 2.1] User preference to rated items (Explicit Feedback)

U/I	Item 1	Item 2	Item 3	...	Item I
User 1		O			O
User 2	O		O		
...		O		O	
User U	O		O		O

[Figure 2.2] User preference with binary labels (Implicit Feedback)

Thesis’s Goal

■ Implicit Feedback으로부터 Matrix Factorization Model 세움

⇩

■ Bayesian Personalized Ranking Learning Algorithm을 통해 loss function이 낮아지는 방향으로 Matrix Factorization에 있는 parameters Update

⇩

■ loss가 가장 낮도록 수렴할 때, Users가 아직 점수를 주지않은 items에 대해 가장 점수예측이 잘 됨

Bayesian Personalized Ranking(BPR)

“To build a personalized ranking function for each user.”

Bayesian Personalized Ranking(BPR)

■ Bayesian Personalized Ranking은 각 User의 personalized ranking을 계산하기 위한 loss function으로, Matrix factorization model에 사용됨(not 알고리즘)

■ 모든 items에 대한 올바른 personalized ranking을 찾기 위한 Bayesian 공식을 사용하여 아래의 probability를 최대화 시킴

$$p(\Theta|>_u) \propto p(>_u|\Theta)\,p(\Theta)$$

Θ는 matrix factorization model의 parameter vector를 나타냄

■ BPR Optimization

$$\begin{aligned}
\text{BPR-OPT} &:= \ln p(\Theta|>_u) \\
&= \ln p(>_u|\Theta)\,p(\Theta) \\
&= \ln \prod_{(u,i,j)\in D_S} \sigma(\hat{x}_{uij})\,p(\Theta) \\
&= \sum_{(u,i,j)\in D_S} \ln \sigma(\hat{x}_{uij}) + \ln p(\Theta) \\
&= \sum_{(u,i,j)\in D_S} \ln \sigma(\hat{x}_{uij}) - \lambda_\Theta ||\Theta||^2
\end{aligned}$$

공식이 Simplified 된 것을 의미

BPR Learning Algorithm

- The gradient of BPR-OPT with respect to the model parameters is:

$$\frac{\partial \text{BPR-Opt}}{\partial \Theta} = \sum_{(u,i,j) \in D_S} \frac{\partial}{\partial \Theta} \ln \sigma(\hat{x}_{uij}) - \lambda_\Theta \frac{\partial}{\partial \Theta} ||\Theta||^2$$

$$\propto \sum_{(u,i,j) \in D_S} \frac{-e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} - \lambda_\Theta \Theta$$

- The model parameters are updated with the learning rate :

$$\Theta \leftarrow \Theta - \alpha \frac{\partial \text{BPR-Opt}}{\partial \Theta} \Longrightarrow \Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)$$

BPR Learning Algorithm

- The gradient of BPR-OPT with respect to the model parameters is:

$$\frac{\partial \text{BPR-OPT}}{\partial \Theta} = \sum_{(u,i,j)\in D_S} \frac{\partial}{\partial \Theta} \ln \sigma(\hat{x}_{uij}) - \lambda_\Theta \frac{\partial}{\partial \Theta} ||\Theta||^2$$

$$\propto \sum_{(u,i,j)\in D_S} \frac{-e^{-\hat{x}_{uij}}}{1+e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} - \lambda_\Theta \Theta$$

- The model parameters are updated with the learning rate :

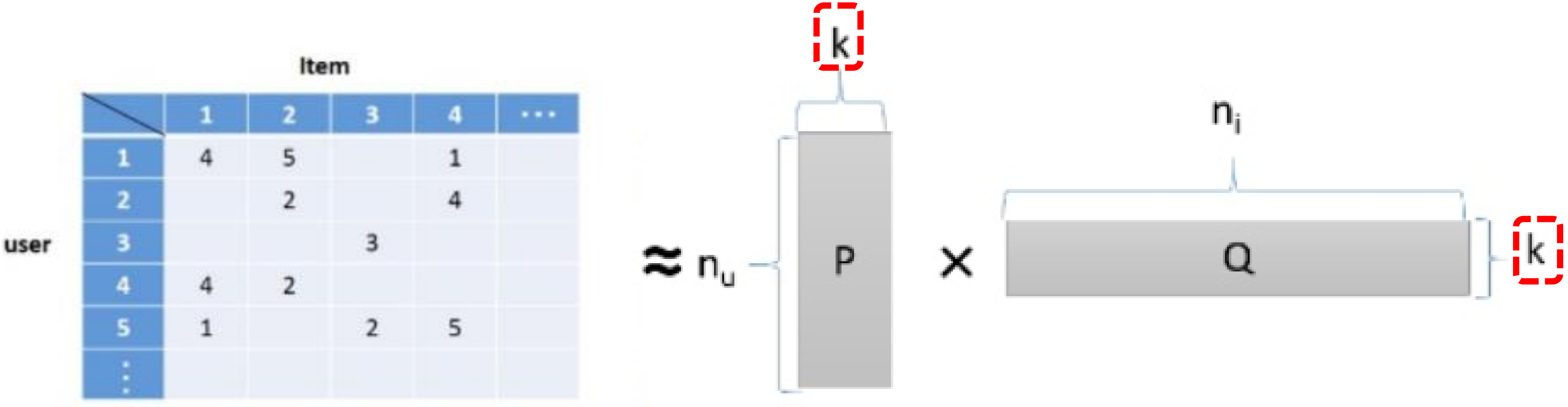
$$\Theta \leftarrow \Theta - \alpha \frac{\partial \text{BPR-OPT}}{\partial \Theta} \Longrightarrow \Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1+e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)$$

Regularization
모든 특징을 사용하나 특징 Θ에 대한 parameter를 줄임

Matrix Factorization(MF)

- 데이터로부터 잠재적인 특징을 찾기 위해 user-based, item-based(or more)의 collaborative 방법을 사용하는 Model

- Factorize a matrix into a product of matrices having k-latent factor



user \ Item	1	2	3	4	…
1	4	5		1	
2		2		4	
3			3		
4	4	2			
5	1		2	5	
⋮					

Rating Matrix User Matrix Item Matrix

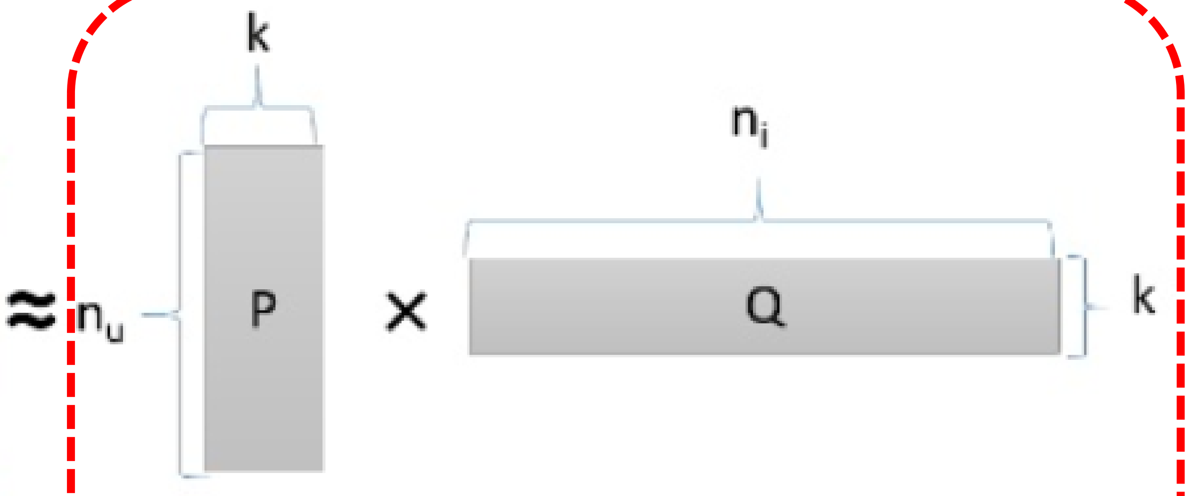
Matrix Factorization(MF)

■ 데이터로부터 잠재적인 특징을 찾기 위해 user-based, item-based(or more)의 collaborative 방법을 사용하는 Model

■ Factorize a matrix into a product of matrices having k-latent factor

user \ Item	1	2	3	4	…
1	4	5		1	
2		2		4	
3			3		
4	4	2			
5	1		2	5	
⋮					

Rating Matrix



User Matrix

Item Matrix

Matrix Factorization(MF)

■ (Basic idea)MF는 어떤 item에 있어서 user의 숨겨진 선호를 예측 할 수 있는데, user-item-pair을 $\hat{x}_{ul} = (u, l)$라 표현함

■ (Proposed idea)논문에서 제안하는 collaborative한 방법으로, triples data를 사용하여 user의 선호를 예측함

$$\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}$$

■ $\hat{x}_{ul}$을 예측하기 위해 matrix factorization하여 $\hat{X} := WH^t$ 를 구함
($W : |U| \times k$ and $H : |I| \times k$)

$$\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}$$

w_{uf} : W matrix의 row로, user의 feature vector h_{if} : H matrix의 row로, item의 feature vector

Example

Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

A matrix of user/item ratings

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

W

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

H

Formula

$$\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}$$

$$\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}$$

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}$$

$$\Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)$$

Example

Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

A matrix of user/item ratings

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

W

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

H

Formula

1. $$\boxed{\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}}$$

2. $$\boxed{\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}}$$

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}$$

$$\Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)$$

Example

Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

A matrix of user/item ratings

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

W

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

H

Formula

1. $$\boxed{\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}}$$

2. $$\boxed{\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}}$$

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}$$

$$\Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)$$

Example

Item

User	W	X	Y	Z
A		4.5	2.0	
B	4.0		3.5	
C		5.0		2.0
D		3.5	4.0	1.0

A matrix of user/item ratings

=

Update

W

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

X

H

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

Formula

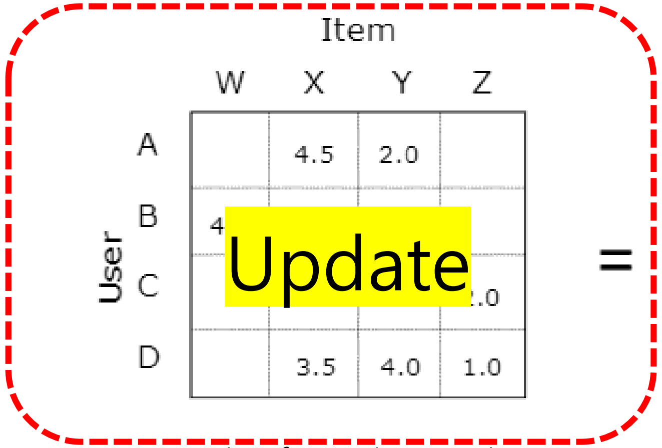
1. $$\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}$$

2. $$\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}$$

3. $$\boxed{\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}}$$

$$\boxed{\Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)}$$

Example



A matrix of user/item ratings

=

A	1.2	0.8
B	1.4	0.9
C	1.5	1.0
D	1.2	0.8

W

X

W	X	Y	Z
1.5	1.2	1.0	0.8
1.7	0.6	1.1	0.4

H

Formula

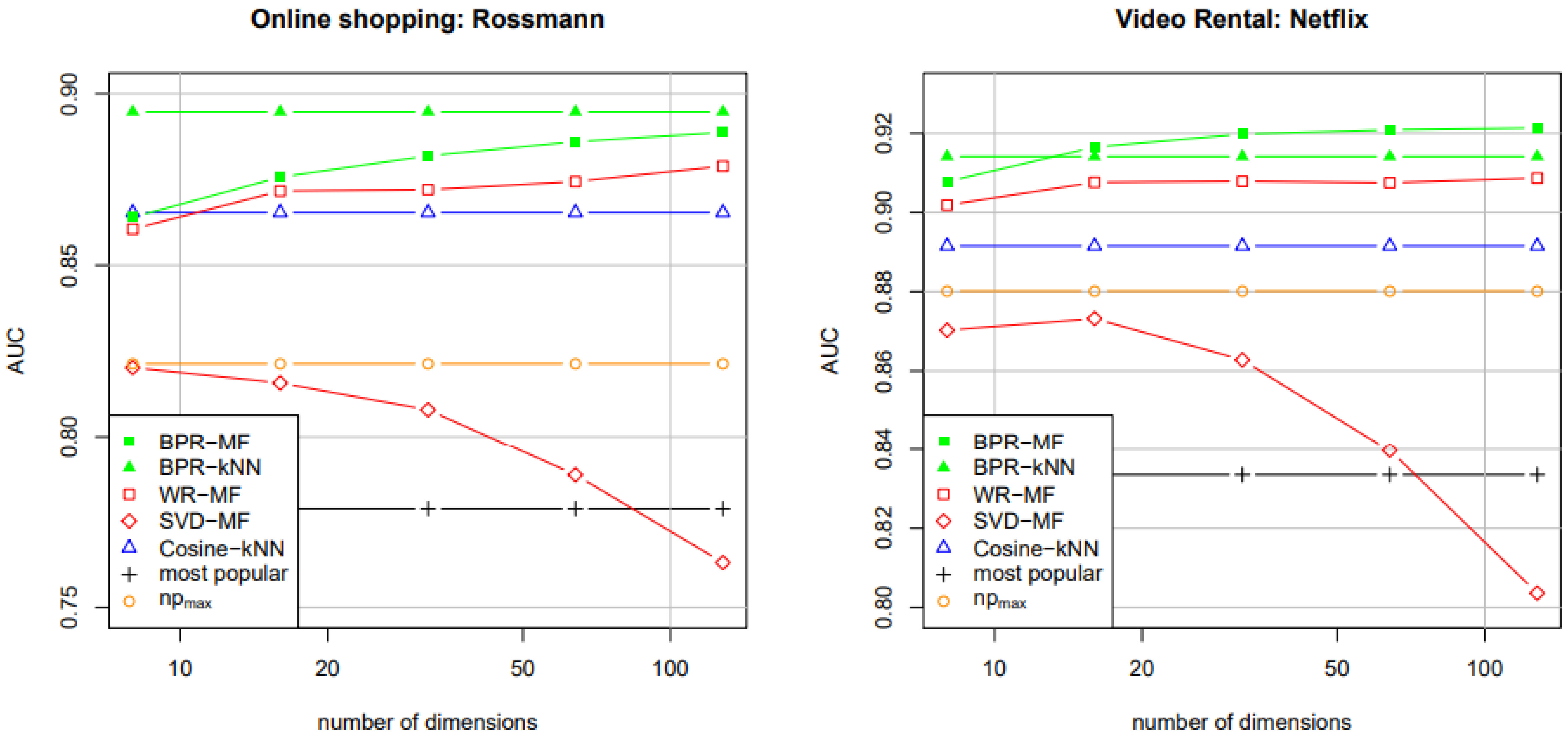
1. $\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj}$

2. $\hat{x}_{ui} = \langle w_u, h_i \rangle = \sum_{f=1}^{k} w_{uf} \cdot h_{if}$

3. $$\boxed{\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} (h_{if} - h_{jf}) & \text{if } \theta = w_{uf}, \\ w_{uf} & \text{if } \theta = h_{if}, \\ -w_{uf} & \text{if } \theta = h_{jf}, \\ 0 & \text{else} \end{cases}}$$

$$\boxed{\Theta \leftarrow \Theta + \alpha \left(\frac{e^{-\hat{x}_{uij}}}{1 + e^{-\hat{x}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{x}_{uij} + \lambda_\Theta \Theta \right)}$$

Results



Online shopping: Rossmann
AUC
0.75
0.80
0.85
0.90
10
20
50
100
number of dimensions
BPR−MF
BPR−kNN
WR−MF
SVD−MF
Cosine−kNN
most popular
np$_{max}$
Video Rental: Netflix
AUC
0.80
0.82
0.84
0.86
0.88
0.90
0.92
10
20
50
100
number of dimensions
BPR−MF
BPR−kNN
WR−MF
SVD−MF
Cosine−kNN
most popular
np$_{max}$

Conclusion

BPR은 Loss function으로, 특정 Model에 의존하지는 않지만 Matrix Factorization이라는 모델에 적용하여 Optimization을 하면 personalized ranking task에 가장 적합한 선택이 됨.