BPR: Bayesian Personalized Ranking from Implicit Feedback

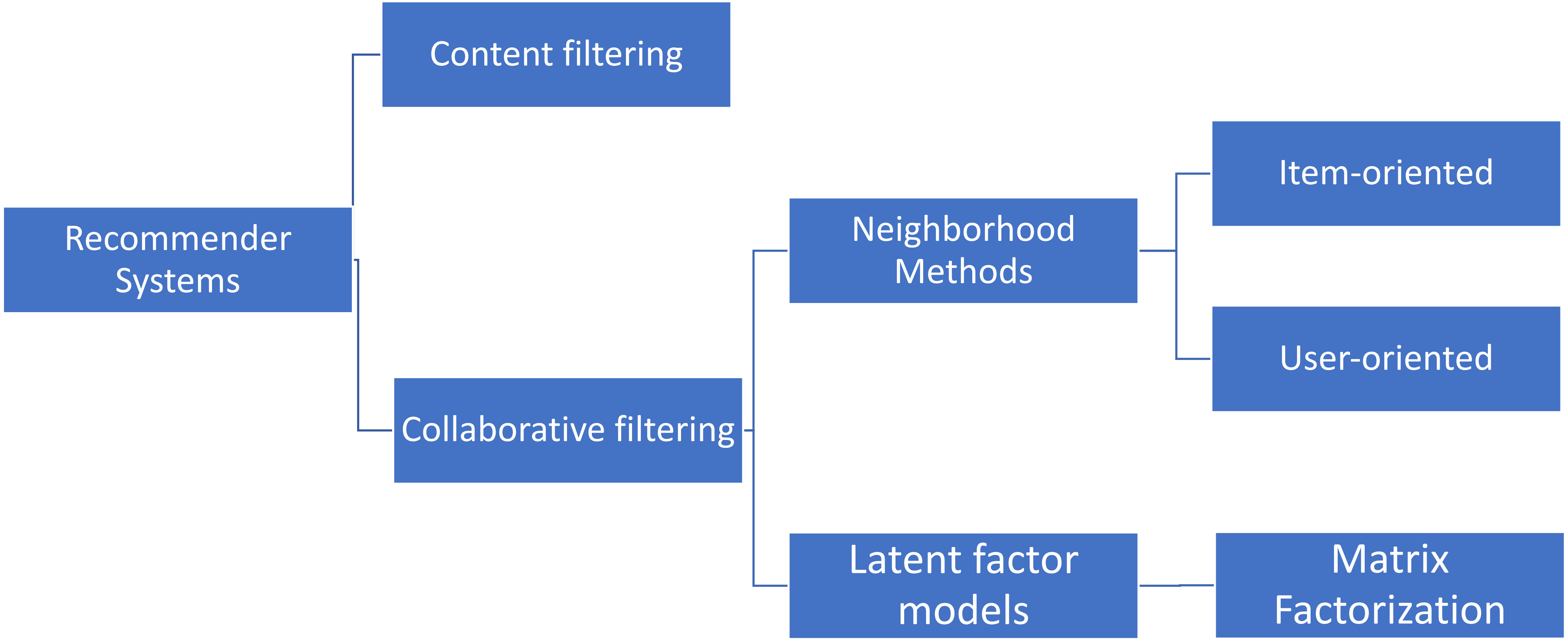
by
Steffen Rendle, Christoph Freudenthaler, Zeno Gantner and Lars Schmidt-Thieme
< Machine Learning Lab, University of Hildsheim >

presented by Sukwon Yun

Contents

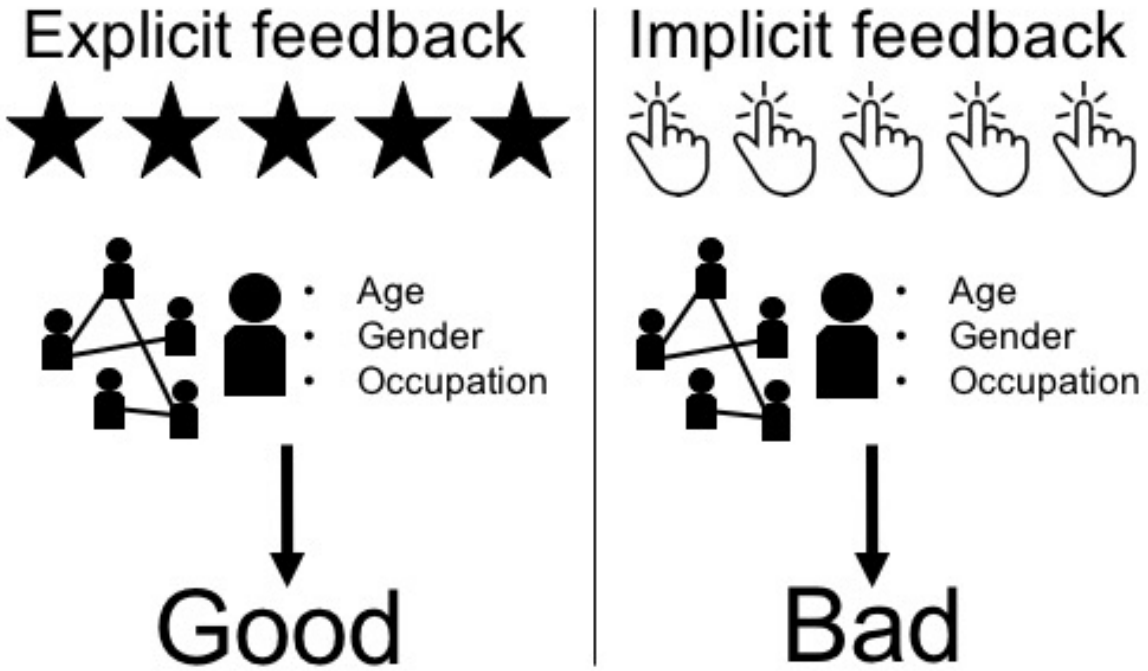
1. Background

- Flow Chart



1. Background

- Implicit feedback



1. Background

- Optimization Method

Minimize sum-of-squared-errors

$$\min_{q^*,p^*} \sum_{(u,i)\in\kappa} (r_{ui} - q_i^T p_u)^2 + \lambda(\| q_i \|^2 + \| p_u \|^2)$$

$$\min_{p^*,q^*,b^*} \sum_{(u,i)\in\kappa} c_{ui}(r_{ui} - \mu - b_u - b_i - p_u^T q_i)^2 + \lambda (\| p_u \|^2 + \| q_i \|^2 + b_u^2 + b_i^2)$$

Maximize log-posterior probability

$$\ln p(U,V|R,\sigma^2,\sigma_V^2,\sigma_U^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij} - U_i^T V_j)^2 - \frac{1}{2\sigma_U^2}\sum_{i=1}^{N} U_i^T U_i - \frac{1}{2\sigma_V^2}\sum_{j=1}^{M} V_j^T V_j$$
$$-\frac{1}{2}\left(\left(\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\right)\ln\sigma^2 + ND\ln\sigma_U^2 + MD\ln\sigma_V^2\right) + C$$

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

1. Background

- Bayesian Probability

"From experienced past, For the interested future"

Likelihood

Prior

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Posterior

2. Introduction

- Personalized Ranking

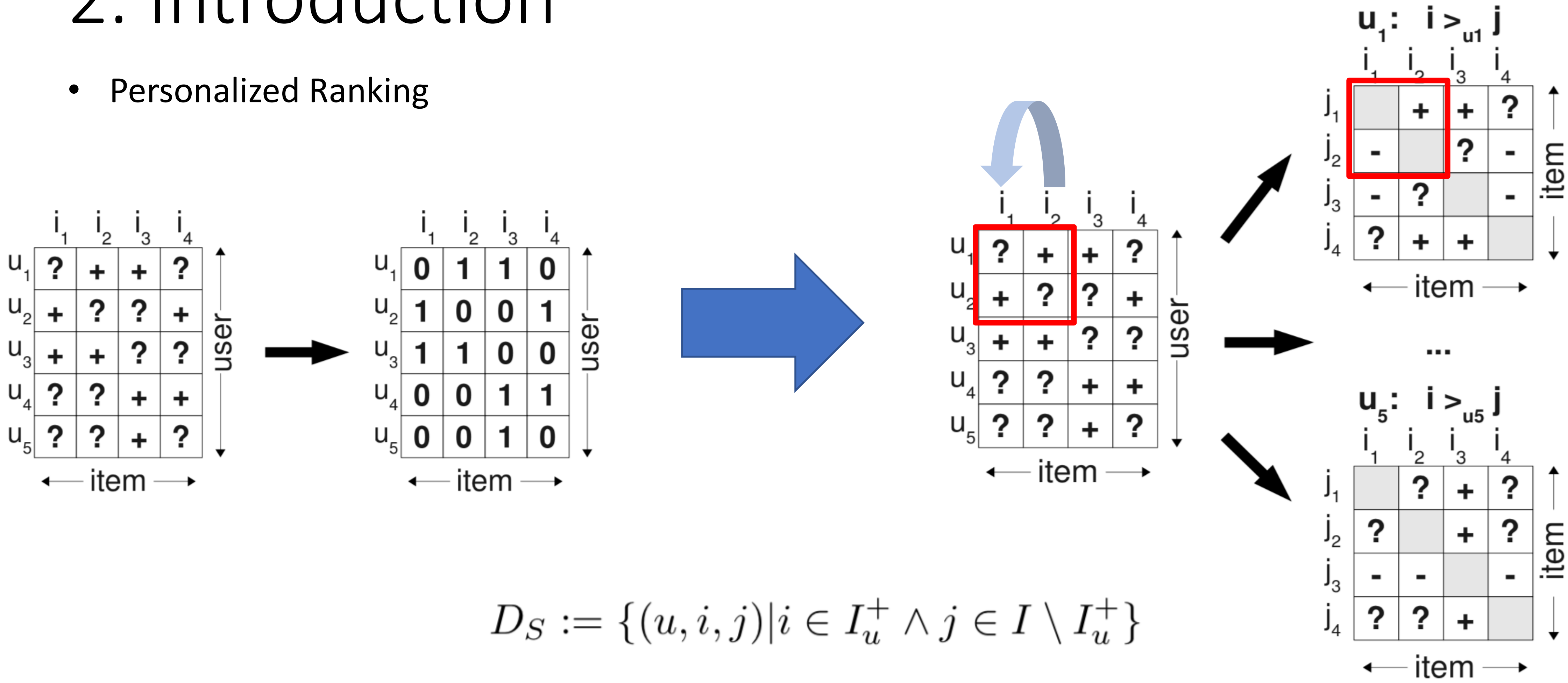
$$S \subseteq U \times I$$

$$>_u \subset I^2$$

$$\begin{aligned}
&\forall i, j \in I : i \neq j \Rightarrow i >_u j \vee j >_u i && (totality)\\
&\forall i, j \in I : i >_u j \wedge j >_u i \Rightarrow i = j && (antisymmetry)\\
&\forall i, j, k \in I : i >_u j \wedge j >_u k \Rightarrow i >_u k && (transitivity)
\end{aligned}$$

2. Introduction

- Personalized Ranking



$$D_S := \{(u,i,j) | i \in I_u^+ \wedge j \in I \setminus I_u^+\}$$

2. Introduction

- **BPR-OPT**

$$\prod_{u \in U} p(>_u |\Theta) = \prod_{(u,i,j) \in U \times I \times I} p(i >_u j|\Theta)^{\delta((u,i,j) \in D_S)} \cdot (1 - p(i >_u j|\Theta))^{\delta((u,j,i) \notin D_S)}$$

$$\delta(b) := \begin{cases} 1 & \text{if } b \text{ is true,} \\ 0 & \text{else} \end{cases}$$

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

$$\prod_{u \in U} p(>_u |\Theta) = \prod_{(u,i,j) \in D_S} p(i >_u j|\Theta)$$

$$\boxed{p(i >_u j|\Theta) := \sigma(\hat{x}_{uij}(\Theta))}$$

2. Introduction

- **BPR-OPT**

$$\boxed{p(i >_u j|\Theta) := \sigma(\hat{x}_{uij}(\Theta))}$$

$$p(\Theta) \sim N(\mathbf{0}, \lambda_\Theta I)$$

$$N(\Theta|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(\Theta - \mu)^T \Sigma^{-1}(\Theta - \mu))$$

$$\propto exp(-\frac{1}{2}\Theta^T(\frac{1}{\lambda_\Theta}I)\Theta)$$

$$= exp(-\frac{1}{2\lambda_\Theta}\Theta^T\Theta) \simeq exp(-\lambda_\Theta\|\Theta\|^2)$$

$$\boxed{p(\Theta) \simeq exp(-\lambda_\Theta\|\Theta\|^2)}$$

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

" Maximize Posterior Probability "

2. Introduction

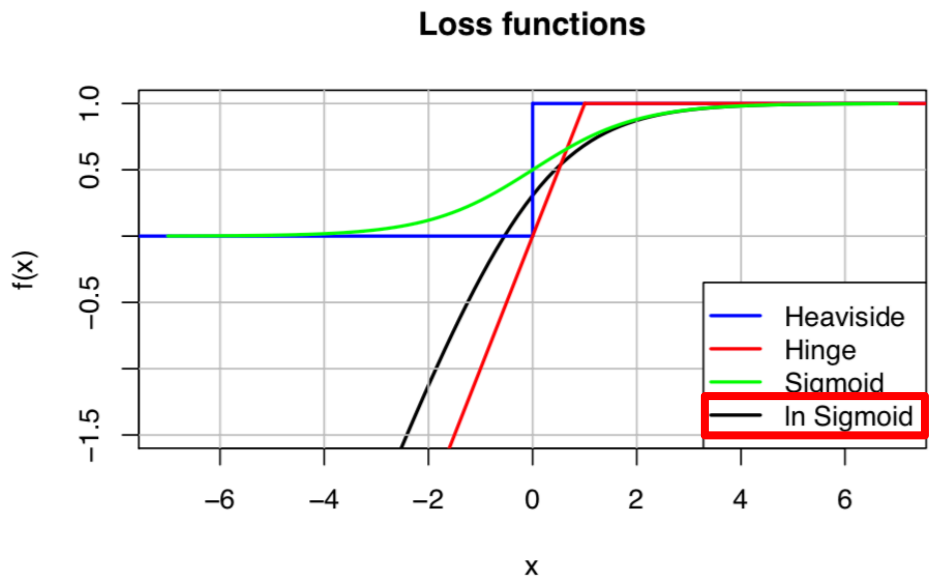
- Analogies between AUC

$$\text{AUC}(u) = \sum_{(u,i,j)\in D_S} z_u \, \delta(\hat{x}_{uij} > 0)$$

$$z_u = \frac{1}{|U|\,|I_u^+|\,|I \setminus I_u^+|}$$

$$\delta(x > 0) = H(x) := \begin{cases} 1, & x > 0 \\ 0, & \text{else} \end{cases}$$

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



2. Introduction

- **LEARNBPR**

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

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

(1) Full gradient descent

$$\Theta \leftarrow \Theta - \alpha \frac{\partial \text{BPR-OPT}}{\partial \Theta}$$

(-) slow convergence
(-) skewness
(-) possibility of i's domination

(2) Stochastic gradient descent with bootstrap

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

(+) faster convergence
(+) less probability in consecutive updating
(+) abandon full cycle

2. Introduction

- Application in Matrix Factorization

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

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

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

2. Introduction

- Application in Adaptive K-Nearest-Neighbor

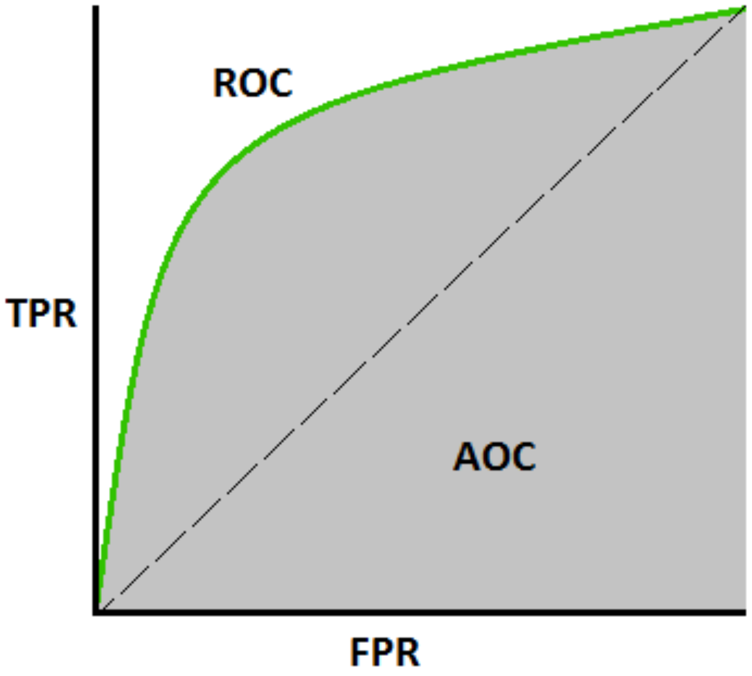
$$\hat{x}_{ui} = \sum_{l \in I_u^+ \wedge l \neq i} c_{il}$$

$$c_{i,j}^{\text{cosine}} := \frac{|U_i^+ \cap U_j^+|}{\sqrt{|U_i^+| \cdot |U_j^+|}}$$

$$\frac{\partial}{\partial \theta} \hat{x}_{uij} = \begin{cases} +1 & \text{if } \theta \in \{c_{il}, c_{li}\} \wedge l \in I_u^+ \wedge l \neq i, \\ -1 & \text{if } \theta \in \{c_{jl}, c_{lj}\} \wedge l \in I_u^+ \wedge l \neq j, \\ 0 & \text{else} \end{cases}$$

2. Introduction

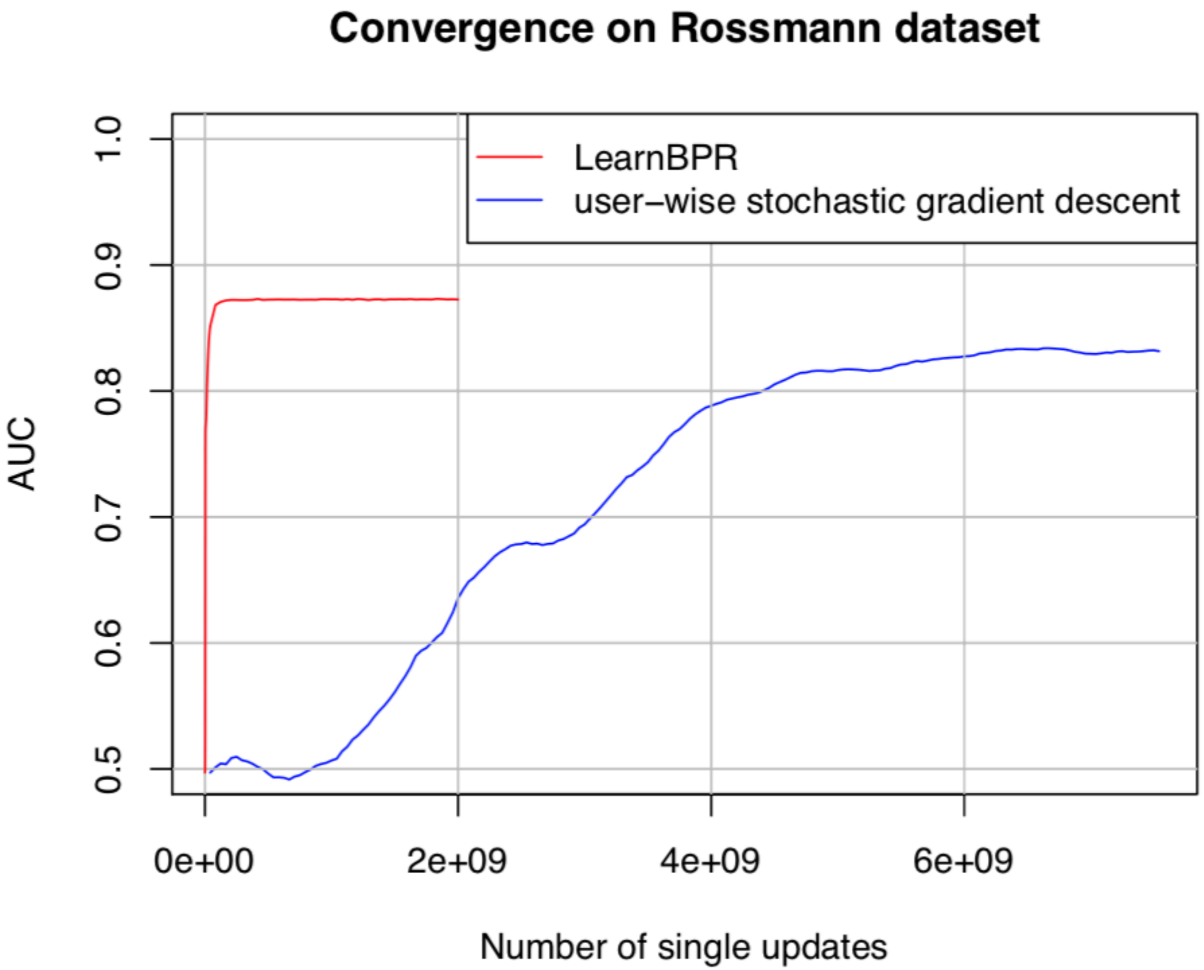
- Computing AUC



$$\mathrm{AUC} = \frac{1}{|U|} \sum_{u} \frac{1}{|E(u)|} \sum_{(i,j) \in E(u)} \delta(\hat{x}_{ui} > \hat{x}_{uj})$$

$$E(u) := \{(i,j) | (u,i) \in S_{\text{test}} \wedge (u,j) \notin (S_{\text{test}} \cup S_{\text{train}})\}$$

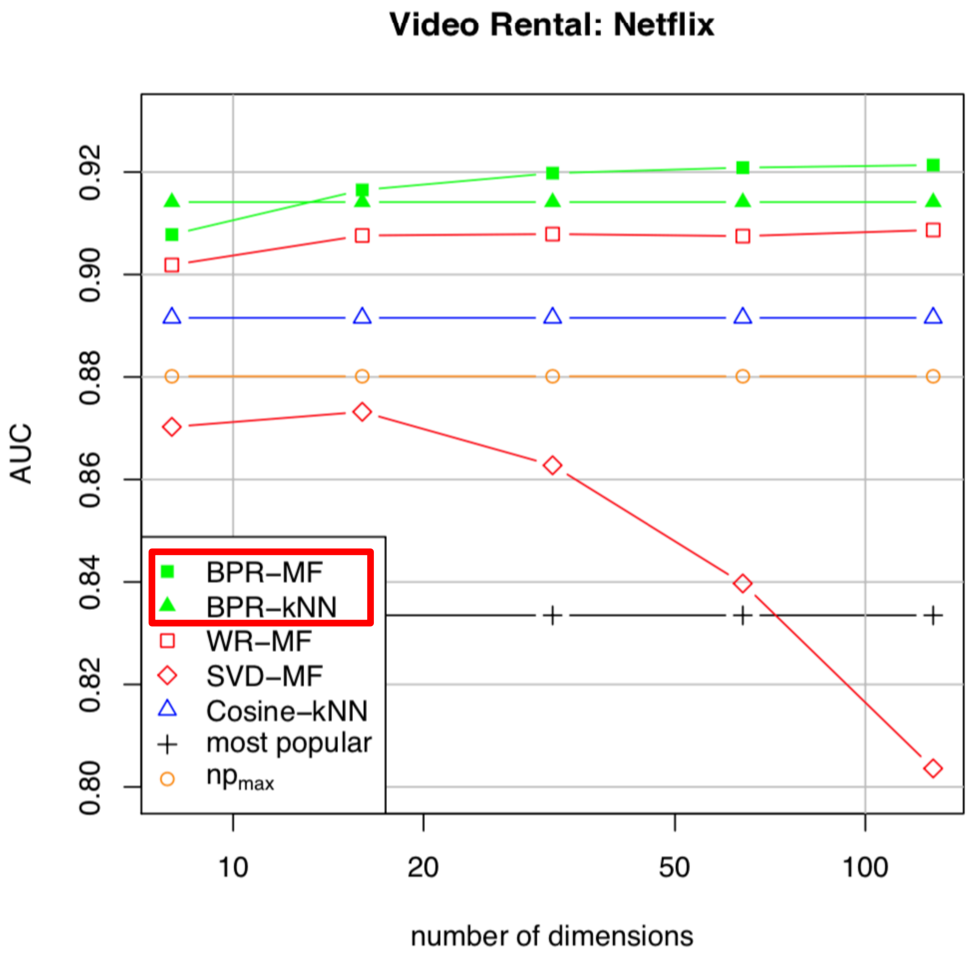
3. Results



Convergence on Rossmann dataset
LearnBPR
user–wise stochastic gradient descent
AUC
1.0
0.9
0.8
0.7
0.6
0.5
0e+00
2e+09
4e+09
6e+09
Number of single updates

3. Results

Online shopping: Rossmann
AUC
0.90
0.85
0.80
0.75
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
0.92
0.90
0.88
0.86
0.84
0.82
0.80
10
20
50
100
number of dimensions
BPR–MF
BPR–kNN
WR–MF
SVD–MF
Cosine–kNN
most popular
np_{max}

4. Implementation

4. Implementation

5. Key takeaways

(1) BPR for personalized ranking

(2) BPR-OPT by Maximizing posterior estimator

(3) LEARN BPR by SGD + bootstrap sampling

(4) Application in MF, kNN

"Importance of optimizing models for the right Criterion"

5. Discussions

- (-) Based on Offline learning (Unable to handle real-time feedback)
- (-) Assumed Independency for user's act
- (-) Loss of information due to focusing on binary cases
  -> Kim, D., & Lee, E. R. (2017). Modified Bayesian personalized ranking for non-binary implicit feedback
- (-) Uniformly selected triples (u, i, j)
  -> Gantner, Z., Drumond, L. (2012, June). Personalized ranking for non-uniformly sampled items.
- (-) AUC isn't the optimal metric

References

(1) Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*.

(2) Mnih, A., & Salakhutdinov, R. R. (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems* (pp. 1257-1264)

(3) Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 426-434)

(4) Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *42*(8), 30-37

(5) Kim, D., & Lee, E. R. (2017). Modified Bayesian personalized ranking for non-binary implicit feedback. *The Korean Journal of Applied Statistics*, *30*(6), 1015-1025.

(6) Gantner, Z., Drumond, L., Freudenthaler, C., & Schmidt-Thieme, L. (2012, June). Personalized ranking for non-uniformly sampled items. In *Proceedings of KDD Cup 2011* (pp. 231-247).