Bayesian Personalized Feature Interaction Selection for Factorization Machines

Yifan Chen, Pengjie Ren, Yang Wang, **Maarten de Rijke**

The main author

- Main author: **Yifan Chen**
- Defended PhD thesis at the University of Amsterdam on October 8, 2018
- Now with NUDT, Changsha, China
- yfchen@nudt.edu.cn

The paper

- Yifan Chen, Pengjie Ren, Yang Wang, and Maarten de Rijke. Bayesian Personalized Feature Interaction Selection for Factorization Machines. In *SIGIR 2019: 42nd international ACM SIGIR conference on Research and Development in Information Retrieval*, page 665–674. ACM, July 2019

- https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/chen-2019-bayesian.pdf

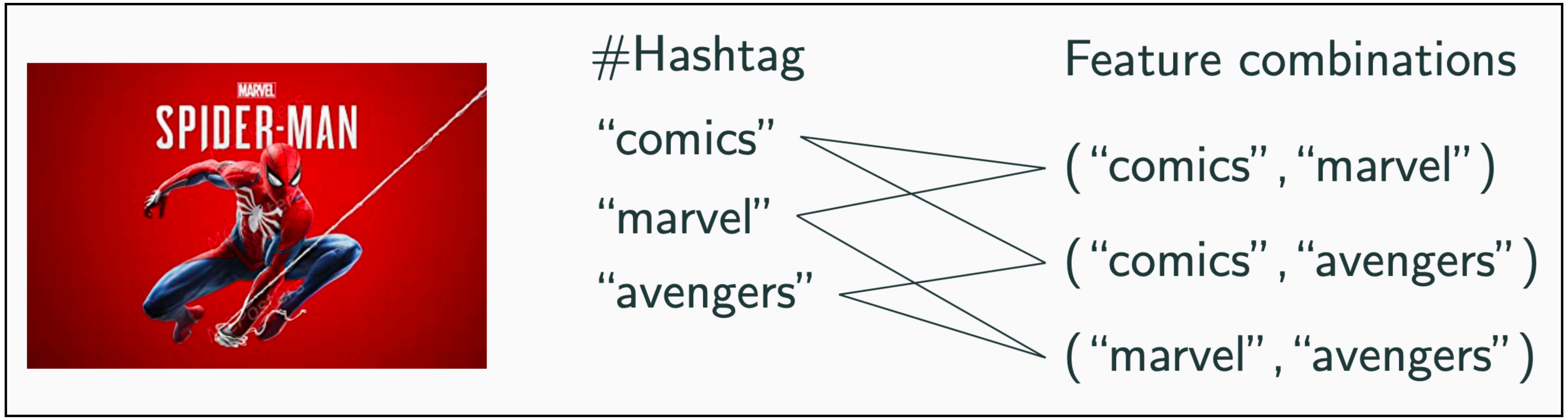
SIGIR 2019

The main points

- We study **personalized feature interaction selection** for factorization machines
- We propose a **Bayesian** personalized feature interaction selection method based on Bayesian variable selection
- [We design an efficient optimization algorithm based on Stochastic Gradient Variational Bayes]

Some details

- Factorization machines:
    - Generic supervised learning method
        - Used for classification and regression
        - Account for feature interactions with factored parameters
        - Usually trained using SGD, ALS, MCMC

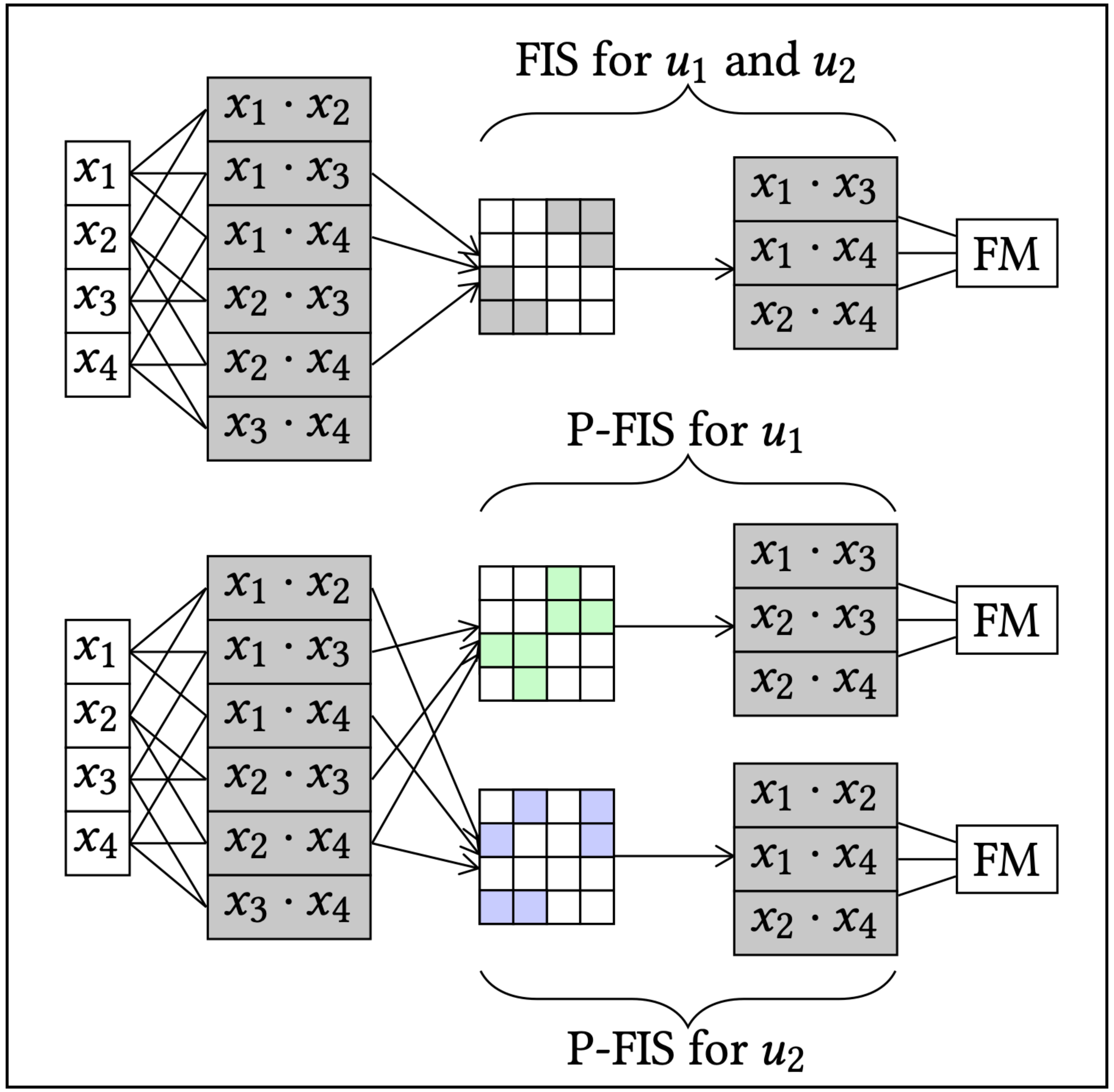


Not all interactions matter to all

- Effective use of historical interactions between users and item
  - Incorporate additional information associated with users or items
  - **High-dimensional** feature space
    - #feature = #user + #item + #additional
    - Not all features or feature interactions are helpful

Not all interactions matter to all

- x_i are features
- $x_1 \cdot x_2$ are feature interactions
- 4×4 matrices indicate masks for selection of feature interactions
- one size fits all vs personalized selection



Rating prediction

- Bias term
- First-order interactions
- Second-order interactions
- Generic vs personalized

$$\hat{r}(\boldsymbol{x}) = b_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d} \sum_{j=i+1}^{d} w_{ij} \cdot x_i x_j$$

$$\hat{r}(\boldsymbol{x}) = b_u + \sum_{i=1}^{d} w_{ui} x_i + \sum_{i=1}^{d} \sum_{j=i+1}^{d} w_{uij} \cdot x_i x_j$$

Bayesian generation model

- To estimate personalized ratings
- Re-parameterization of interaction weights
- Use hereditary spike-and-slab prior to reduce number of candidate interactions

Algorithm 1 Generation procedure

```
for each user u ∈ 𝒰 do
    for for each feature i ∈ ℱ do
        draw first-order interaction selection variable s_ui ~ Bernoulli(π_1);
        draw first-order interaction weight w̃_i ~ 𝒩(0, 1);
        w_ui = s_ui · w̃_i.
for for each feature pair i, j ∈ ℱ do
    draw second-order interaction selection variable s_uij ~ p(s_uij | s_ui, s_uj);
    draw second-order interaction weight w̃_ij ~ 𝒩(0, 1);
    w_uij = s_uij · w̃_ij.
for for each feature vector x ∈ 𝒳 do
    calculate the rating prediction r̂(x) by Eq. (3);
    draw r(x) ~ p(r | r̂(x)).
```

The results

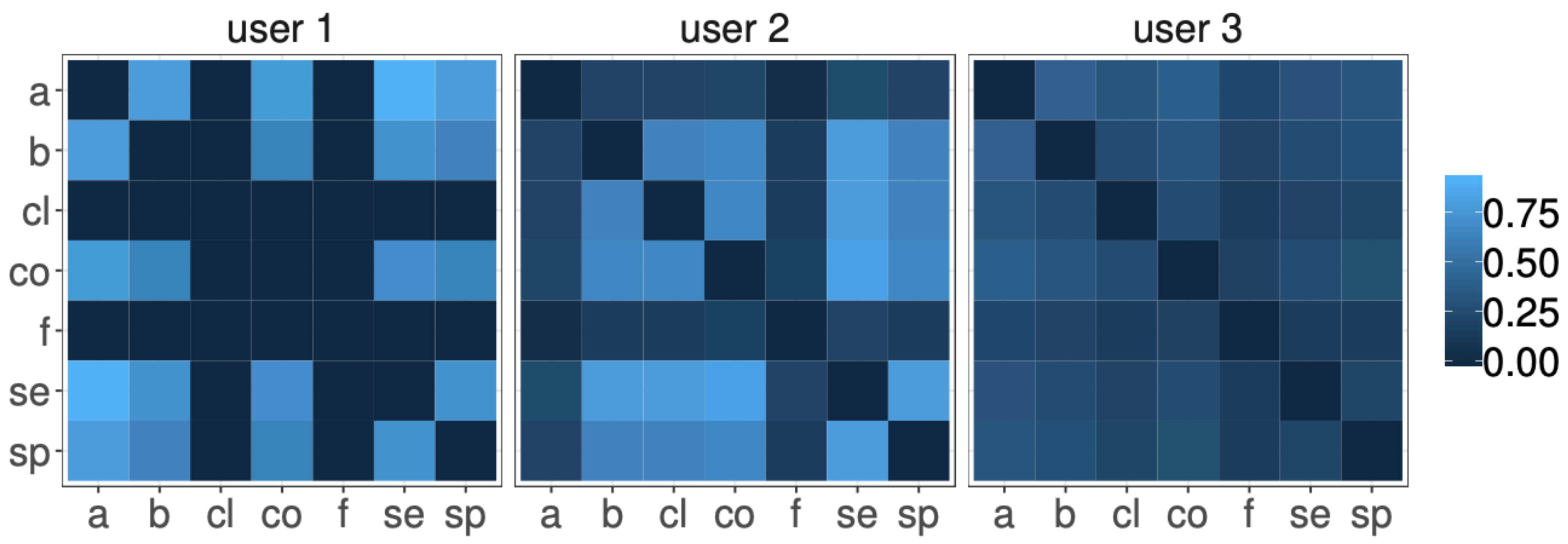
- Significant improvements over linear and non-linear factorization machines
- Multiple datasets

A closer look

- Examples based on MovieLens HetRec dataset
  - MovieLens 10M plus IMDB plus Rotten Tomatoes

	48 Hrs.	Spider-Man	Lethal Weapon	True Lies
action (a)	√	√	√	
buddy (b)	√		√	√
clever (cl)	√			√
comedy (co)	√		√	√
funny (f)		√		√
sequel (se)		√	√	
special effects (sp)		√		√
user 1	–	top-5	top-1	top-10
user 2	–	top-5	top-5	top-10
user 3	–	–	–	–

Ticks indicate whether a movie has the feature. We also indicate for each user whether the movie is within the top-1, top-5 or top-10 recommendation.



What's next?

- Extend method to higher-order interactions or multi-view and multimodal factorizations
- Consider group-level selections of interactions to speed up training

- Paper available at: https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/chen-2019-bayesian.pdf
- Code available at: https://github.com/yifanclifford/BP-FIS