

A Preference-Based Bandit Framework for Personalized Recommendation

Maryam Tavakol and Ulf Brefeld

Paderborn, Nov 8, 2016

Introduction

Personalized Recommendation

```
graph TD; A[Personalized Recommendation] --> B((Multi-armed bandits)); A --> C((Preference Learning));
```

The diagram illustrates the relationship between Personalized Recommendation and two related concepts. A gray rounded rectangle at the top contains the text 'Personalized Recommendation'. Two red arrows originate from the bottom of this rectangle, pointing to two blue ovals below. The left oval is labeled 'Multi-armed bandits' and the right oval is labeled 'Preference Learning'.

Multi-armed bandits

Preference Learning


Recommendation


naketano
BRAVE NEW WORD

Naketano
DARTH III - Hooded Sweatshirt - indigo
★★★★★ (95 customer reviews)

[More Naketano](#) | [More Hoodie](#)

 **FREE
SHIPPING & RETURNS**

 **100 DAYS
RETURN POLICY**

 **FREE HOTLINE
0800 240 10 20**



€ 69.95

incl VAT

Color



Height

Select size



[Size Chart](#)


Available within 1-3 working days

Add to Cart

Add to Wishlist



Recommendation





Naketano
DARTH III - Hooded Sweatshirt - indigo
★★★★★ (95 customer reviews)
More Naketano | More Hoodie

FREE SHIPPING & RETURNS

100 DAYS RETURN POLICY

FREE HOTLINE 0800 240 10 20



Height


Select size

Size Chart

Available within 1-3 working days

Add to Cart

Add to Wishlist



You may also like



Naketano
REORDER II - Sweatshirt - anthracite melange
€ 69.95



Naketano
GREASE RAGS II - Hoodie - dark gray melange
€ 59.95



Naketano
brazzo - Sweat Jacket - dark gray melange
from 69,90 €



Naketano
MANDY - Sweatshirt - indigo blue melange
€ 39.95



Naketano
DARTH III - Hoodie - candy red melange
€ 64.90

Preference Model

- Item i : {*Shirt*, *Blue*, *Women*, *Cheap*}
- Item k : {*Polo shirt*, *White*, *Women*, *Expensive*}

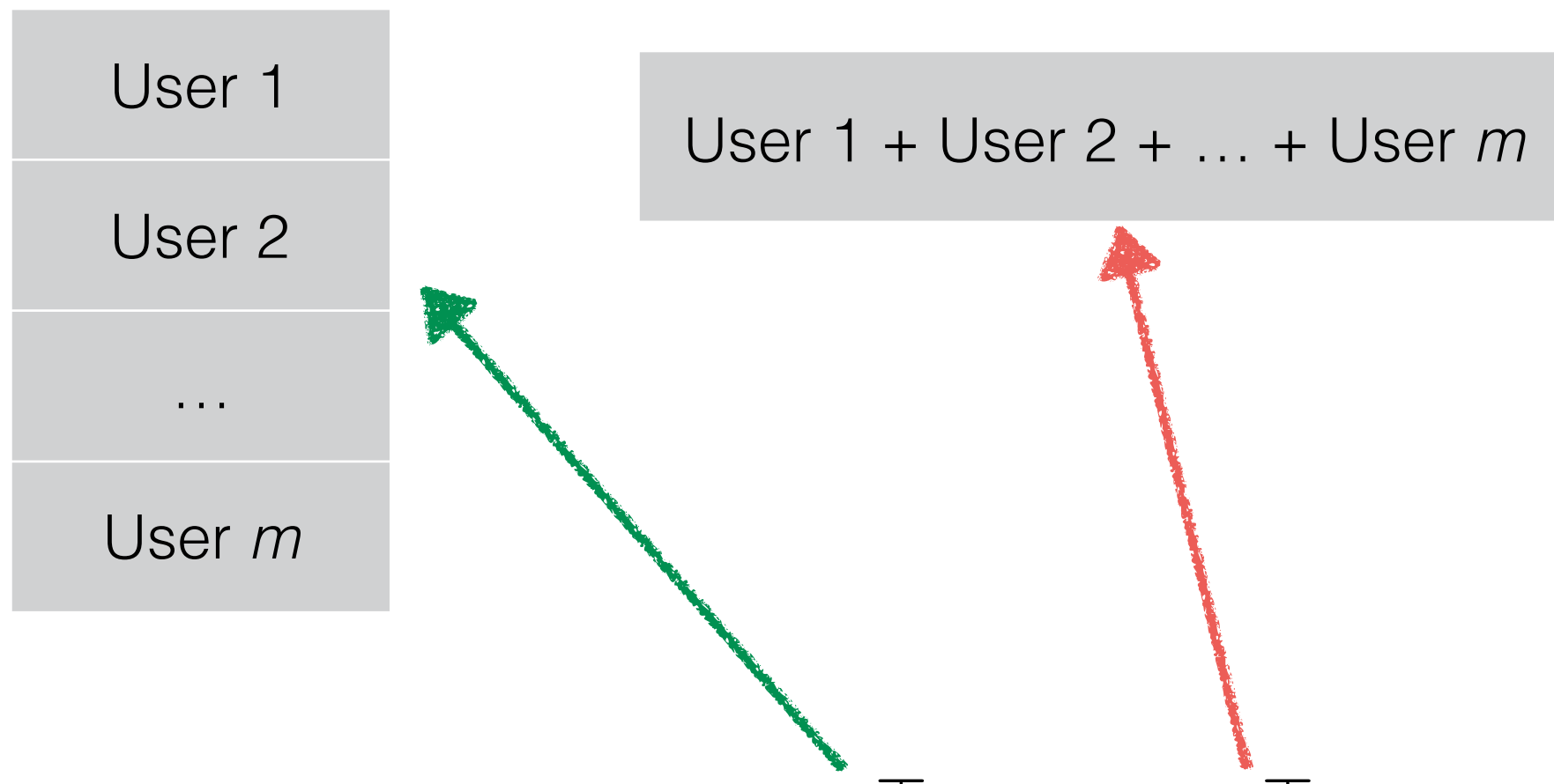
Item i $>$ Item k :

{*Shirt-Polo shirt*, *Blue-White*, *Women-Women*, *Cheap-Expensive*}

$$\mathbf{z}_{i \succ k} := \mathbf{z}_i - \mathbf{z}_k$$

Payoff Model

- **Personalized model** + **average component**



$$\mathbb{E}[r_{t,i \succ k} | u_t = u_j] = \boldsymbol{\beta}_t^\top \mathbf{z}_{i \succ k} + \boldsymbol{\theta}^\top \mathbf{z}_{i \succ k}$$

Personalized Recommendation with Qualitative Bandit

- For **$t = 1, \dots, T$** :
 1. The world generates some context
 2. The learner chooses an action
 3. The world reacts with a reward
- Choosing the arm with the highest **mean reward + confidence interval**

(General case of LinUCB)

Unified Optimization

- Solving the objective function in **dual space**
 - With arbitrary loss function
 - Using Fenchel-Legendre conjugate

$$\begin{aligned} \sup_{\boldsymbol{\alpha}} \quad & -C \sum_{t=1}^T V^*\left(-\frac{\alpha_t}{C}, r_t\right) - \frac{1}{2} \boldsymbol{\alpha}^\top \mathbf{Z} \mathbf{Z}^\top \boldsymbol{\alpha} \\ & - \frac{1}{2\mu} \sum_j \boldsymbol{\alpha}^\top (\mathbf{Z} \circ \boldsymbol{\phi}_j) (\mathbf{Z} \circ \boldsymbol{\phi}_j)^\top \boldsymbol{\alpha}. \end{aligned}$$

Squared Loss

$$\begin{aligned} \max_{\alpha} \quad & -\frac{1}{2C} \alpha^\top \alpha + r^\top \alpha \\ & -\frac{1}{2} \alpha^\top \left[ZZ^\top + \frac{1}{\mu} \left(\sum_j \phi_j \otimes \phi_j^\top \right) \circ ZZ^\top \right] \alpha \end{aligned}$$

- The problem reduces to standard quadratic optimization
- Model parameters (θ, β_j) , are obtained from α

Squared Loss

- In the contextual bandit framework:

- Mean:

$$\boldsymbol{\beta}_t^\top \mathbf{z}_{i \succ k} + \boldsymbol{\theta}^\top \mathbf{z}_{i \succ k}$$

- Confidence bound:

$$c \sqrt{\mathbf{z}_{i \succ k}^\top (Z^\top Z + \lambda I)^{-1} \mathbf{z}_{i \succ k}}$$

Algorithm

```
for  $t = 1, 2, \dots, T$  do  
  Observe the user  $u_j$   
  for all  $\{a_i, a_k\} \in A_t$  do  
    Observe the features  $\mathbf{z}_i$  and  $\mathbf{z}_k$   
     $\mathbf{z}_{i \succ k} := \mathbf{z}_i - \mathbf{z}_k$   
    
$$p_{i,k} = (\boldsymbol{\beta}_j + \boldsymbol{\theta})^\top \mathbf{z}_{i \succ k} + c \sqrt{\mathbf{z}_{i \succ k}^\top (Z^\top Z + \lambda I)^{-1} \mathbf{z}_{i \succ k}}$$
  
  end for  
  Choose arm  $a_t = \arg \max_i p_{i,k}$ , and observe payoff  $r_t$   
  Obtain  $\alpha$  and update  $\boldsymbol{\theta}$  and  $\boldsymbol{\beta}_j$   
end for
```

Summary

- Personalized recommendation
- Pairwise learning in bandit framework
- Optimization in dual space
- Learning algorithm for squared loss

Thanks for your attention

Questions?

Email: tavakol@leuphana.de