LEUPHANA
UNIVERSITÄT LÜNEBURG

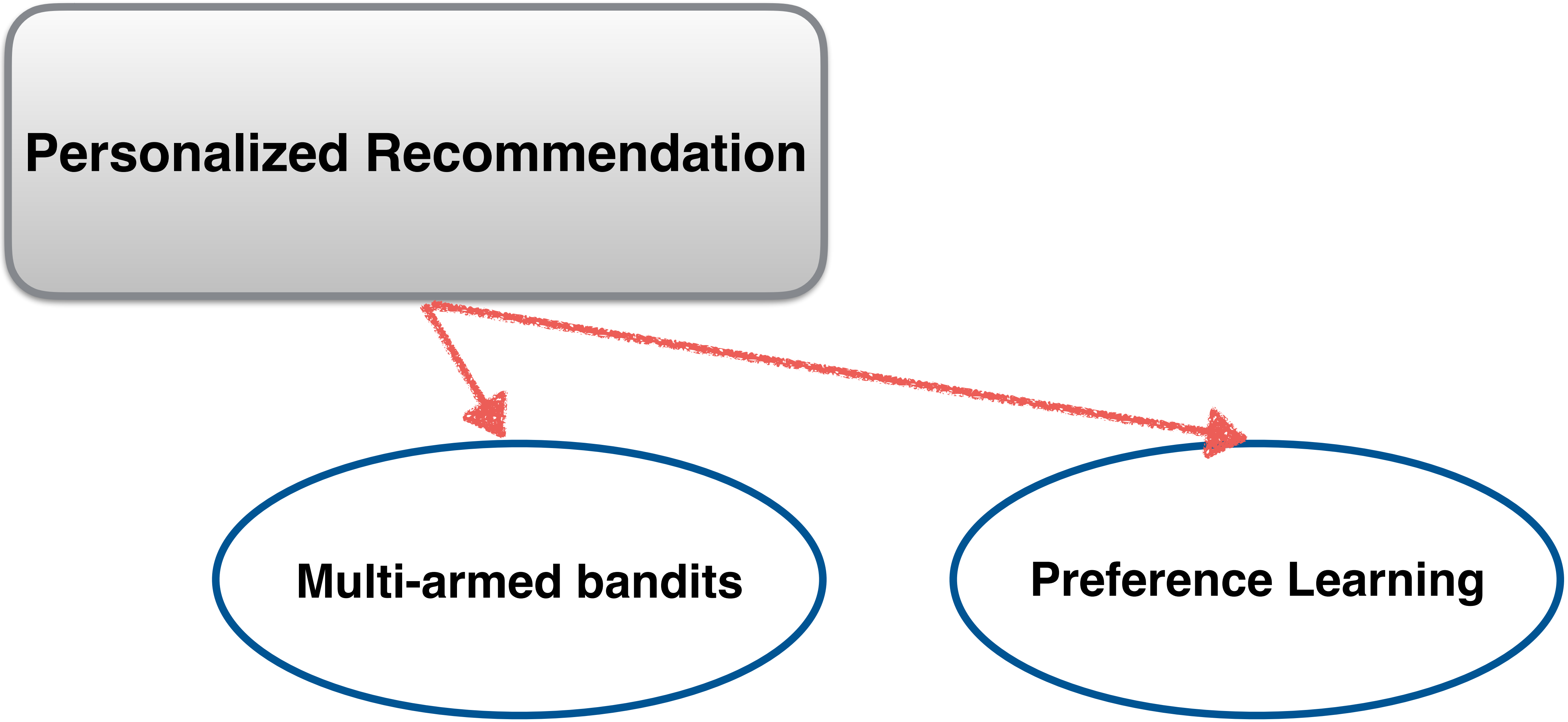
TECHNISCHE
UNIVERSITÄT
DARMSTADT

A Preference-Based Bandit Framework for Personalized Recommendation

Maryam Tavakol and Ulf Brefeld

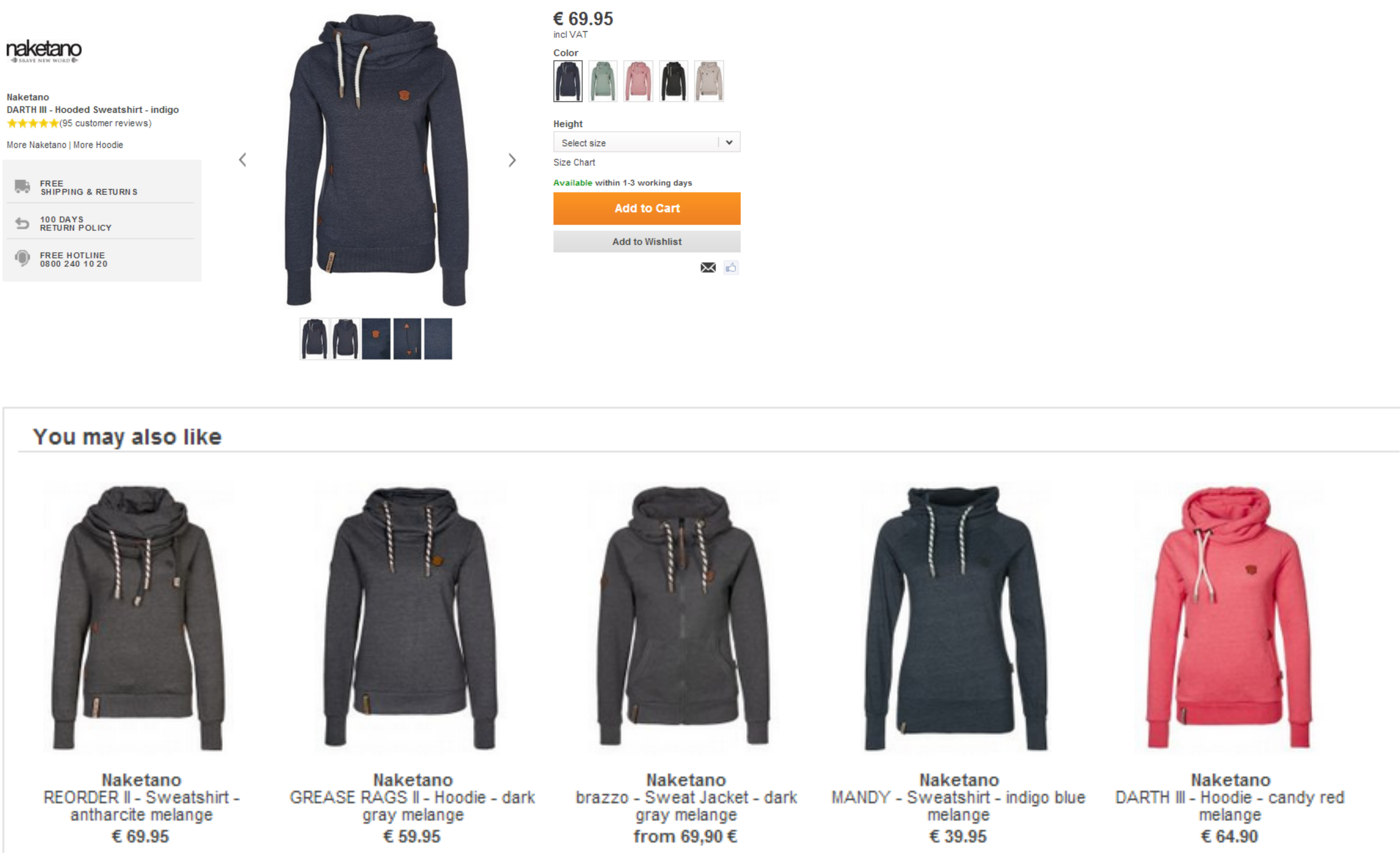
Paderborn, Nov 8, 2016

Introduction



Recommendation

Recommendation



Preference Model

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

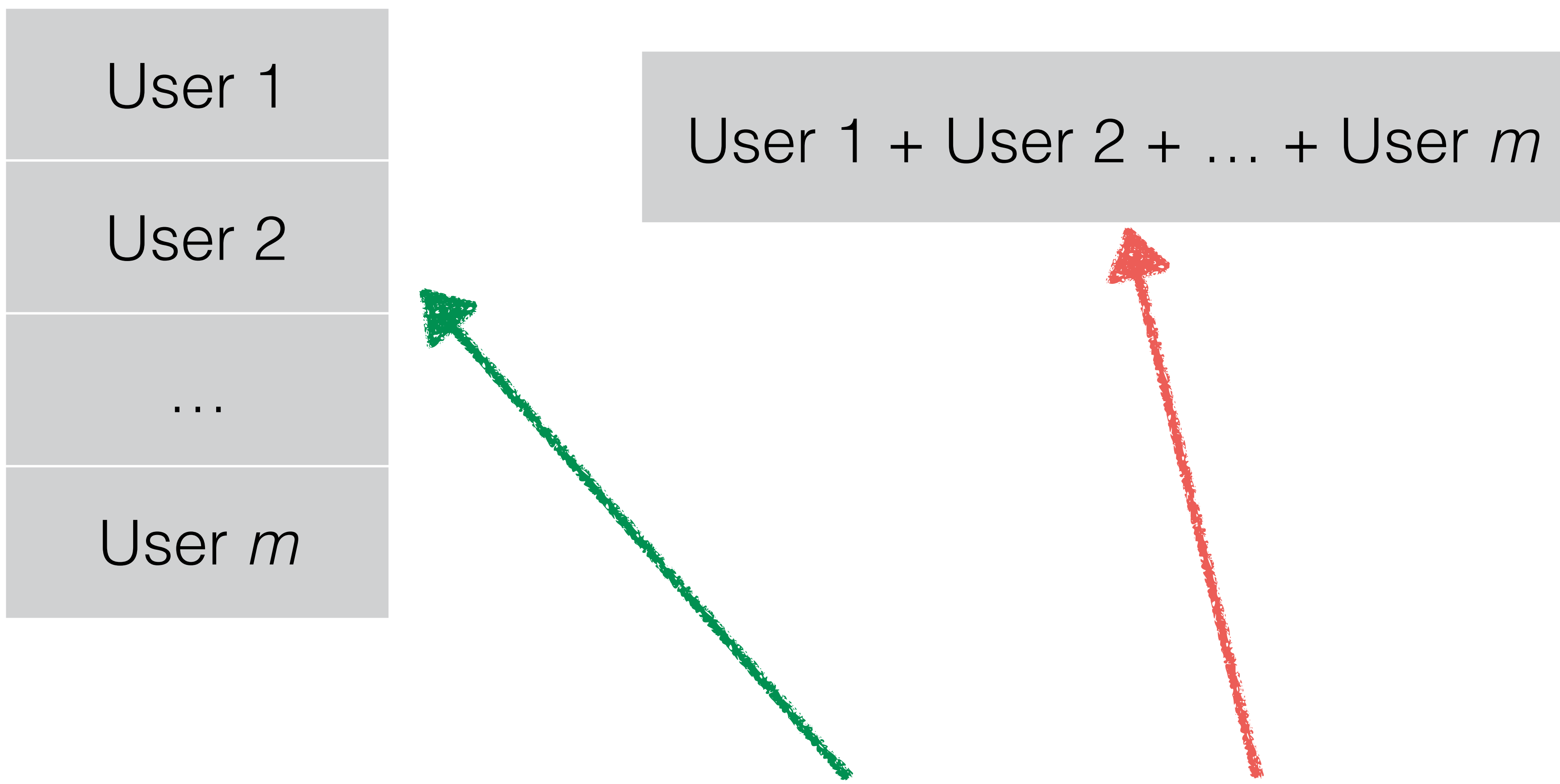
Item *i* ≻ Item *k*:

{*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, …, 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

$$\sup_{\boldsymbol{\alpha}} \quad -C\sum_{t=1}^{T} V^*(-\frac{\alpha_t}{C}, r_t) - \frac{1}{2}\boldsymbol{\alpha}^\top Z Z^\top \boldsymbol{\alpha}$$
$$- \frac{1}{2\mu}\sum_{j} \boldsymbol{\alpha}^\top (Z \circ \boldsymbol{\phi_j})(Z \circ \boldsymbol{\phi_j})^\top \boldsymbol{\alpha}.$$

Squared Loss

$$\max_{\boldsymbol{\alpha}} \quad -\frac{1}{2C}\boldsymbol{\alpha}^\top\boldsymbol{\alpha} + \boldsymbol{r}^\top\boldsymbol{\alpha}$$

$$-\frac{1}{2}\boldsymbol{\alpha}^\top[ZZ^\top + \frac{1}{\mu}(\sum_j \boldsymbol{\phi}_j \otimes \boldsymbol{\phi}_j^\top) \circ ZZ^\top]\boldsymbol{\alpha}$$

- The problem reduces to standard quadratic optimization
- Model parameters ($\boldsymbol{\theta}$, $\boldsymbol{\beta_j}$), are obtained from $\boldsymbol{\alpha}$

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:

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

Algorithm

for $t = 1, 2, ..., T$ **do**

Observe the user u_j

for all $\{a_i, a_k\} \in A_t$ **do**

Observe the features $\mathbf{z}_i$ and $\mathbf{z}_k$

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

$p_{i,k} = (\boldsymbol{\beta}_j + \boldsymbol{\theta})^\top \mathbf{z}_{i \succ k} + c\sqrt{\boldsymbol{z}_{i \succ k}^\top (Z^\top Z + \lambda I)^{-1} \boldsymbol{z}_{i \succ k}}$

end for

Choose arm $a_t = \arg\max_i p_{i,k}$, and observe payoff r_t

Obtain $\boldsymbol{\alpha}$ 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