LEUPHANA
UNIVERSITÄT LÜNEBURG

TECHNISCHE
UNIVERSITÄT
DARMSTADT

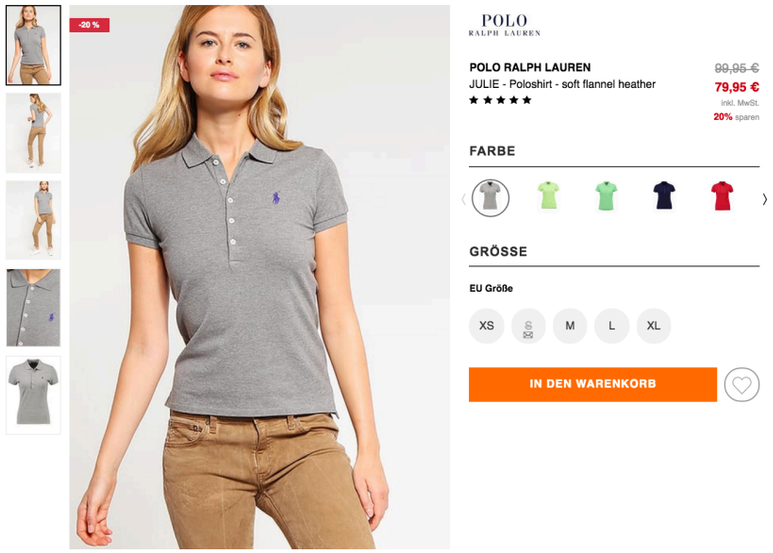
A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations

Maryam Tavakol and **Ulf Brefeld**
{tavakol,brefeld}@leuphana.de

Skopje - Sep 21, 2017

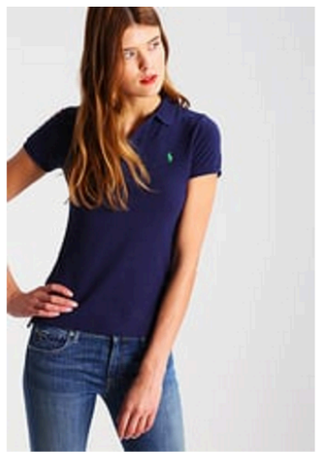
Recommendation

Recommendation



DAS KÖNNTE DIR AUCH GEFALLEN

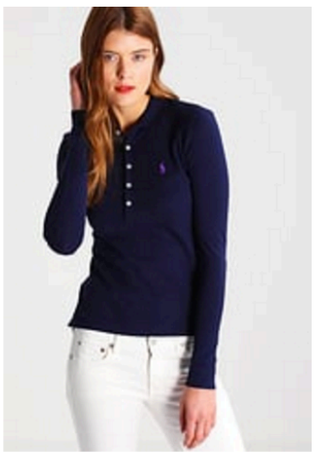
Polo Ralph Lauren
119,95 €



Polo Ralph Lauren
89,95 €

Tommy Hilfiger
59,95 €

Polo Ralph Lauren
~~129,95 €~~ 103,95 €



Polo Ralph Lauren
~~109,95 €~~ 87,95 €

Personalization


The Washington Post

SPORTS

D

THURSDAY, APRIL 19, 2012

SOCCER

United's missed shot
A 1-1 draw against the expansion Montreal Impact disappoints a squad that could've used a victory. D3

ON WASHINGTONPOST.COM/SPORTS

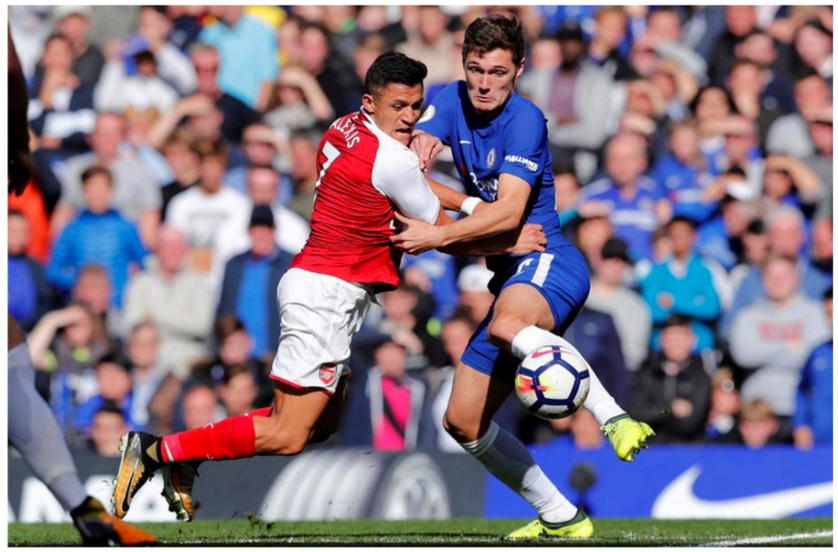
Tracee Hamilton *Today, 11 a.m.* Columnist discusses world of Washington-area sports in a Q&A.

Capitals Insider *Tonight, 7:30 p.m.* Can't get to Game 4? Talk Caps-Bruins on our open thread.

Schedule predictor Cast a vote on which games the Redskins win, and what their record will be.

PRO BASKETBALL

Crawford is Buck wild
Jordan Crawford scores 32 to lead the Wizards past the playoff-chasing Milwaukee Bucks, 125-112. D3



After Chelsea Ties Arsenal, a Feeling Something Has Been Lost

On the diamond, another Nats gem

Short-Term Zeitgeist

iPhone 8 has a low carbon footprint, says Apple VP

Lisa Jackson, head of Apple's
CEO Tim Cook's leadership ha
becoming more green.

JIM WILSON/THE NEW YORK TIMES

STATE OF THE ART

The iPhone 8: A Worthy Refinement Before the Next Generation

Apple's new iPhone 8 and 8 Plus have a ring of familiarity. But the phones may feel like a solid upgrade from older models because of their new processor.

Flying to US to get an iPhone X is cheaper than buying in Europe. It's also illegal

Proposed Approach

- **Goal:** Combination of long- and short-term interests of users in one unified model

  ➡ **Long-term** part + **Short-term** component

✤ **s.t.:** Generality in terms of optimization

- **Framework:** Contextual Multi-Armed Bandit (MAB)
  - e.g., LinUCB

Unified Model

current user

outcome

item features

context

bias term

$$\mathbb{E}[r_{t,a_i}|u_j] = \underbrace{\boldsymbol{\theta}_i^\top \boldsymbol{x}_t}_{Short-term} + \underbrace{\boldsymbol{\beta}_j^\top \boldsymbol{z}_{a_i}}_{Long-term} + b_i$$

parameters of short-term model

parameters of long-term model

General Optimization

- Objective function with arbitrary loss, $V(\cdot, r_t)$

$$\inf_{\substack{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n \\ \boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_m \\ \boldsymbol{b}}} \quad \frac{1}{T}\sum_{t=1}^{T} V(\boldsymbol{\theta}_t^\top \boldsymbol{x}_t + \boldsymbol{\beta}_t^\top \boldsymbol{z}_t + b_t, r_t) + \frac{\lambda}{2}\sum_i \|\boldsymbol{\theta}_i\|^2 + \frac{\hat{\mu}}{2}\sum_j \|\boldsymbol{\beta}_j\|^2$$

Regularization

General Optimization

- Using the Fenchel-Legendre conjugate of loss function in the *dual space*:

$$\sup_{\boldsymbol{\alpha}, \mathbf{1}^\top \boldsymbol{\alpha}=0} \quad -C \sum_{t=1}^{T} V^*(-\frac{\alpha_t}{C}, r_t) - \frac{1}{2} \boldsymbol{\alpha}^\top [(\sum_i \boldsymbol{\delta}_i \otimes \boldsymbol{\delta}_i^\top) \circ XX^\top + \frac{1}{\mu}(\sum_i \boldsymbol{\phi}_i \otimes \boldsymbol{\phi}_i^\top) \circ ZZ^\top] \boldsymbol{\alpha}$$

Kernel trick

Optimization

- Gradient-based approaches (in dual or primal)
  - Calculating the gradient depends on the loss function
- Model parameters, $(\boldsymbol{\theta}_i, \boldsymbol{\beta_j})$, are obtained from $\boldsymbol{\alpha}$
- Kernel functions applicable

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Algorithm

for $t = 1, 2, ..., T$ **do**
  Observe user u_t and context $\mathbf{x}_t$
  for all $a \in A_t$ **do**
    Observe arm features $\mathbf{z}_a$
    $v_{t,a}$ = mean reward + confidence bound
  end for
  Choose arm $a_t = \arg\max_a v_{t,a}$, and observe payoff r_t
  Obtain $\boldsymbol{\alpha}$ by optimizing the objective
  Compute $\boldsymbol{\theta_t}$ and $\boldsymbol{\beta_t}$ from $\boldsymbol{\alpha}$
end for

Instantiation: Squared Loss

- Conjugate of the loss function:

$$V^*(-\frac{\alpha_t}{C}, r_t) = \frac{1}{2C^2}\alpha_t^2 - \frac{1}{C}\alpha_t r_t$$

  - Becomes a standard quadratic optimization with a constraint

- Confidence bound: $c\sqrt{\mathbf{x}_t^\top (X^\top X)^{-1}\mathbf{x}_t + \mathbf{z}_t^\top (Z^\top Z)^{-1}\mathbf{z}_t}$

Instantiation: Logistic Loss

- Conjugate of the loss function:

$$V^*(-\frac{\alpha_t}{r_t}, r_t) = (1 - \frac{\alpha_t}{Cr_t}) \log(1 - \frac{\alpha_t}{Cr_t}) + \frac{\alpha_t}{Cr_t} \log(\frac{\alpha_t}{Cr_t})$$

- Confidence bound:

Diagonal matrix of sigmoid model

$$c\sqrt{\boldsymbol{x}_t^\top (X^\top V_a X)^{-1} \boldsymbol{x}_t + \boldsymbol{z}_t^\top (Z^\top V_u Z)^{-1} \boldsymbol{z}_t}$$

Model Simplification

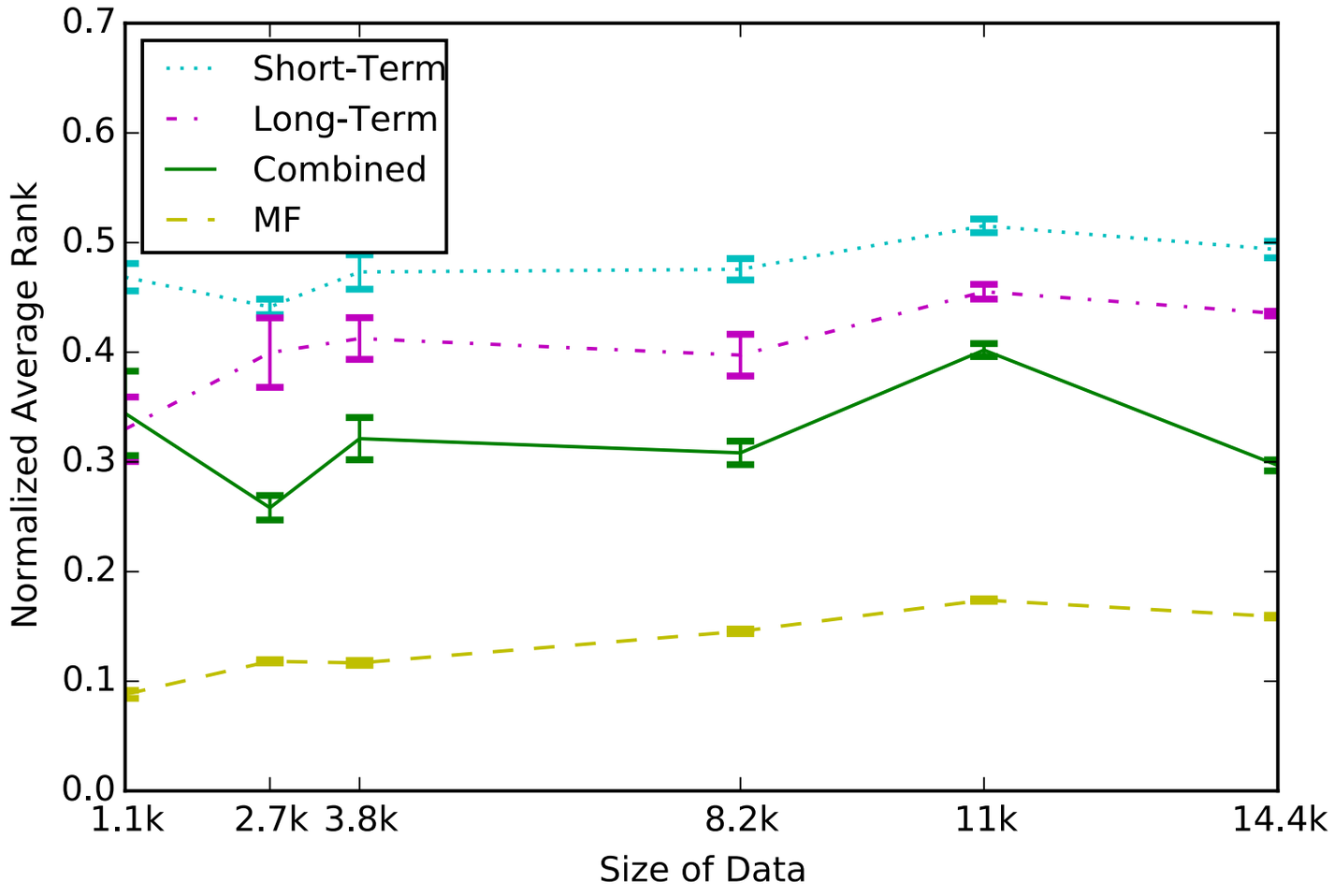
- Focus on the item model
  - **Short-Term:** $\mathbb{E}[r_{t,a_i}] = \boldsymbol{\theta}_i^\top \boldsymbol{x}_t$
  - **Short-Term+Average:** $\mathbb{E}[r_{t,a_i}] = \boldsymbol{\theta}_i^\top \boldsymbol{x}_t + \boldsymbol{\beta}^\top \boldsymbol{z}_{a_i}$
- Focus on the user model
  - **Long-Term:** $\mathbb{E}[r_{t,a_i}|u_j] = \boldsymbol{\beta}_j^\top \boldsymbol{z}_{a_i}$
  - **Long-Term+Average:** $\mathbb{E}[r_{t,a_i}|u_j] = \boldsymbol{\beta}_j^\top \boldsymbol{z}_{a_i} + \boldsymbol{\theta}^\top \boldsymbol{z}_{a_i}$

Empirical Study

- Using **squared** loss function
- Dataset: User transactions from Zalando*
- Baseline: Matrix Factorization (MF)
- Performance measure: normalized average rank

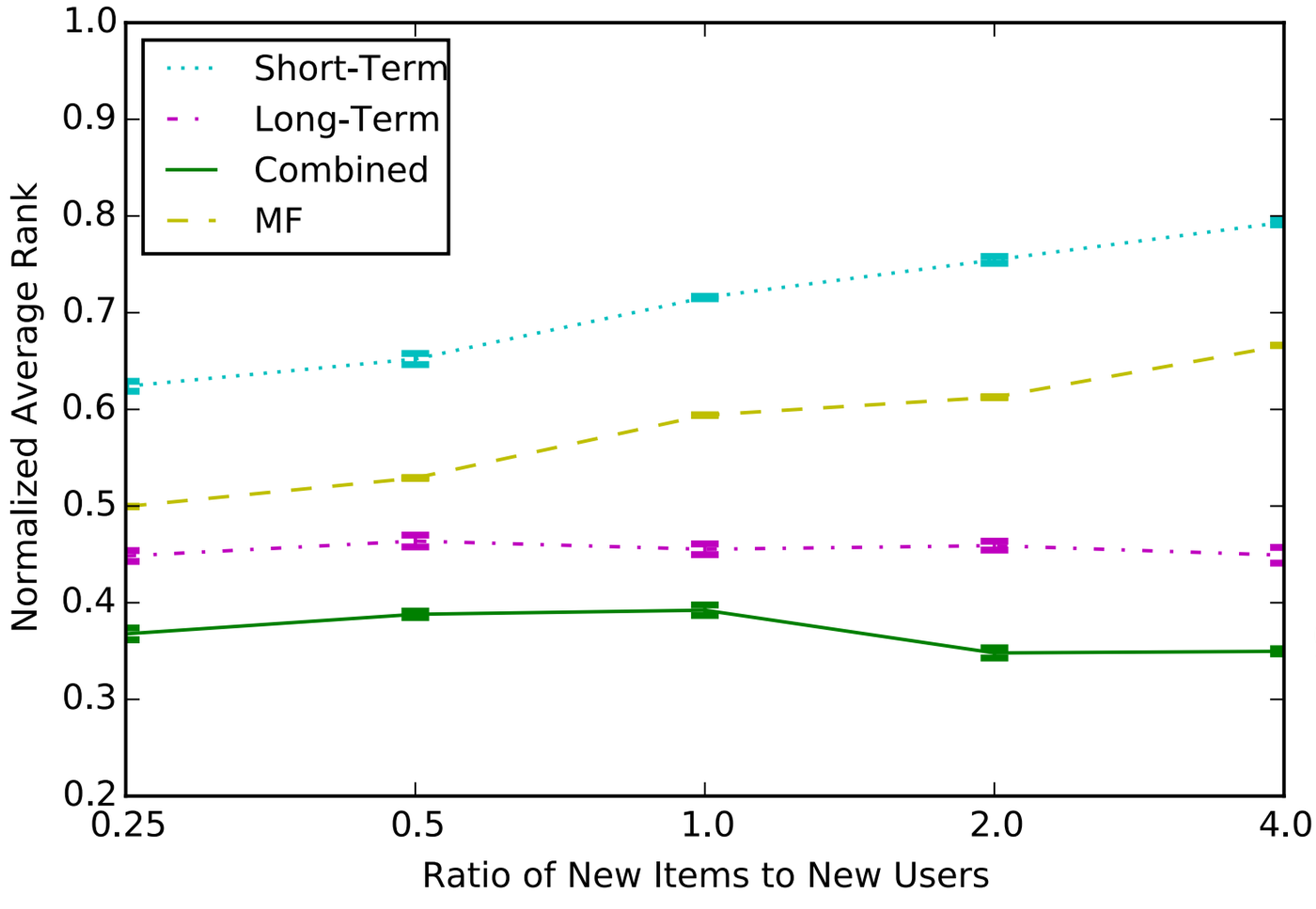
No New User/Item

- The combined approach outperforms either short- or long-term models —but not the baseline!



Cold Start Scenarios

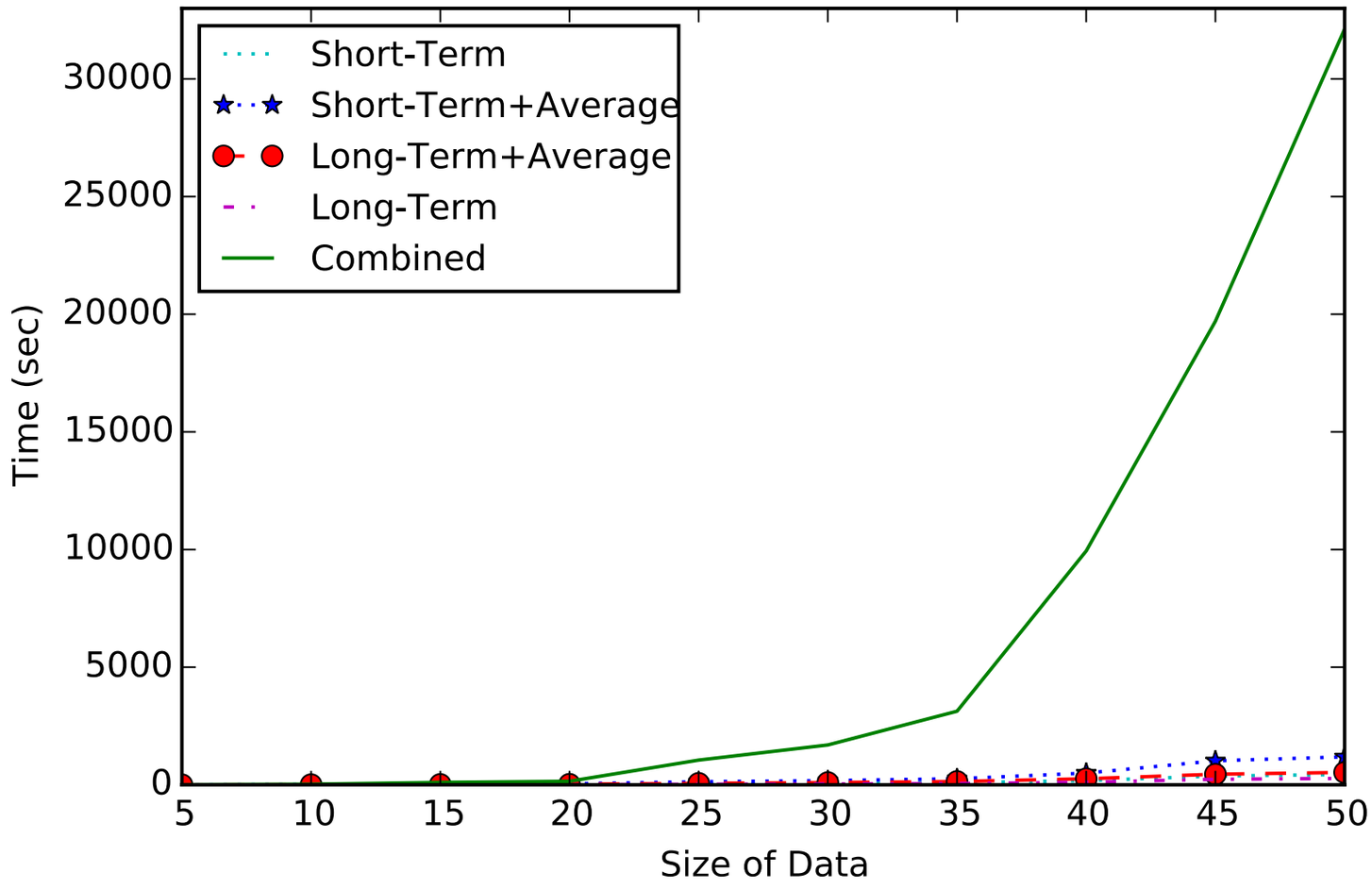
- Robustness of combined model in case of new user/item



generalizes well for both cases

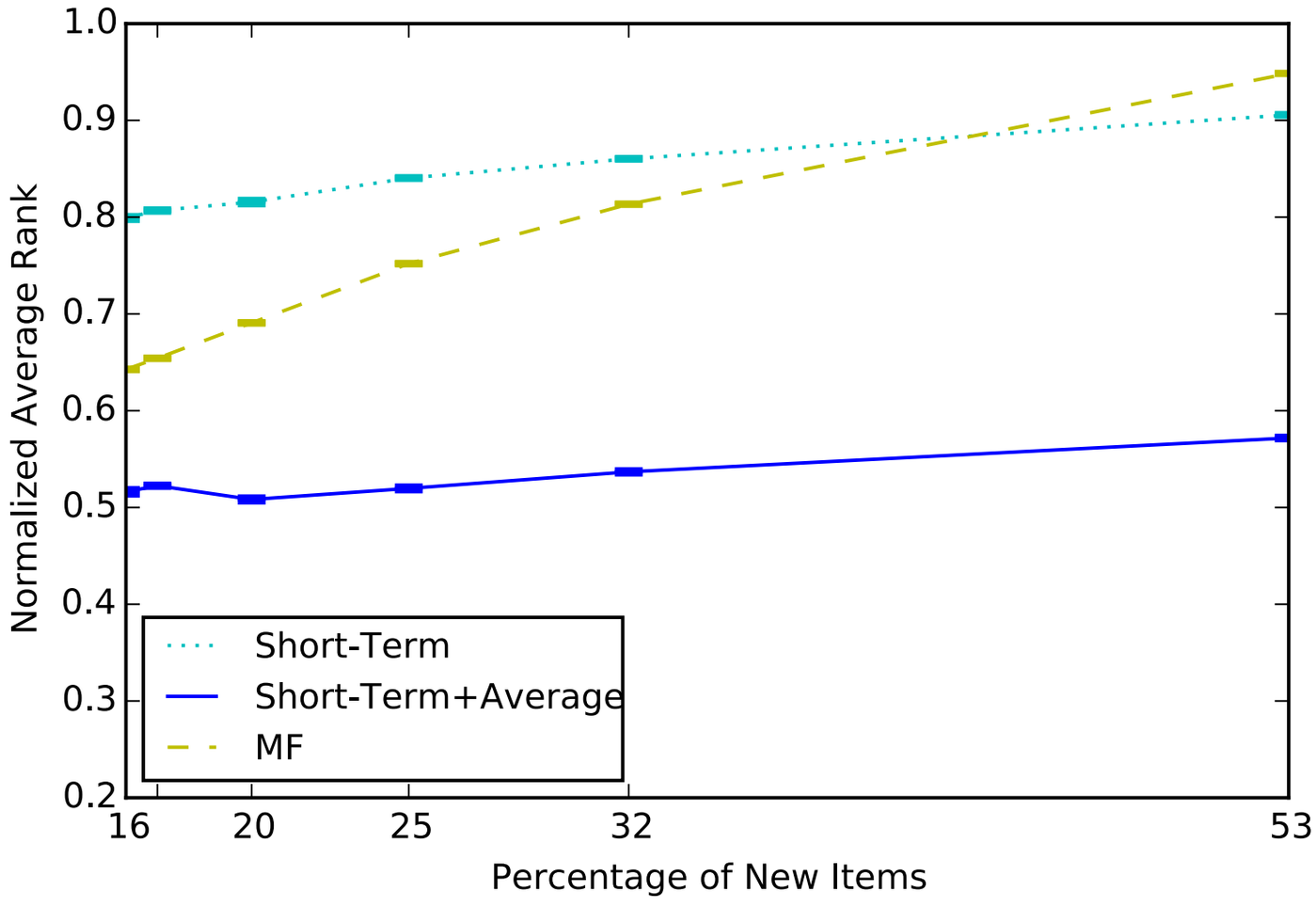
Time Complexity

- The optimization time in combined model is exponential



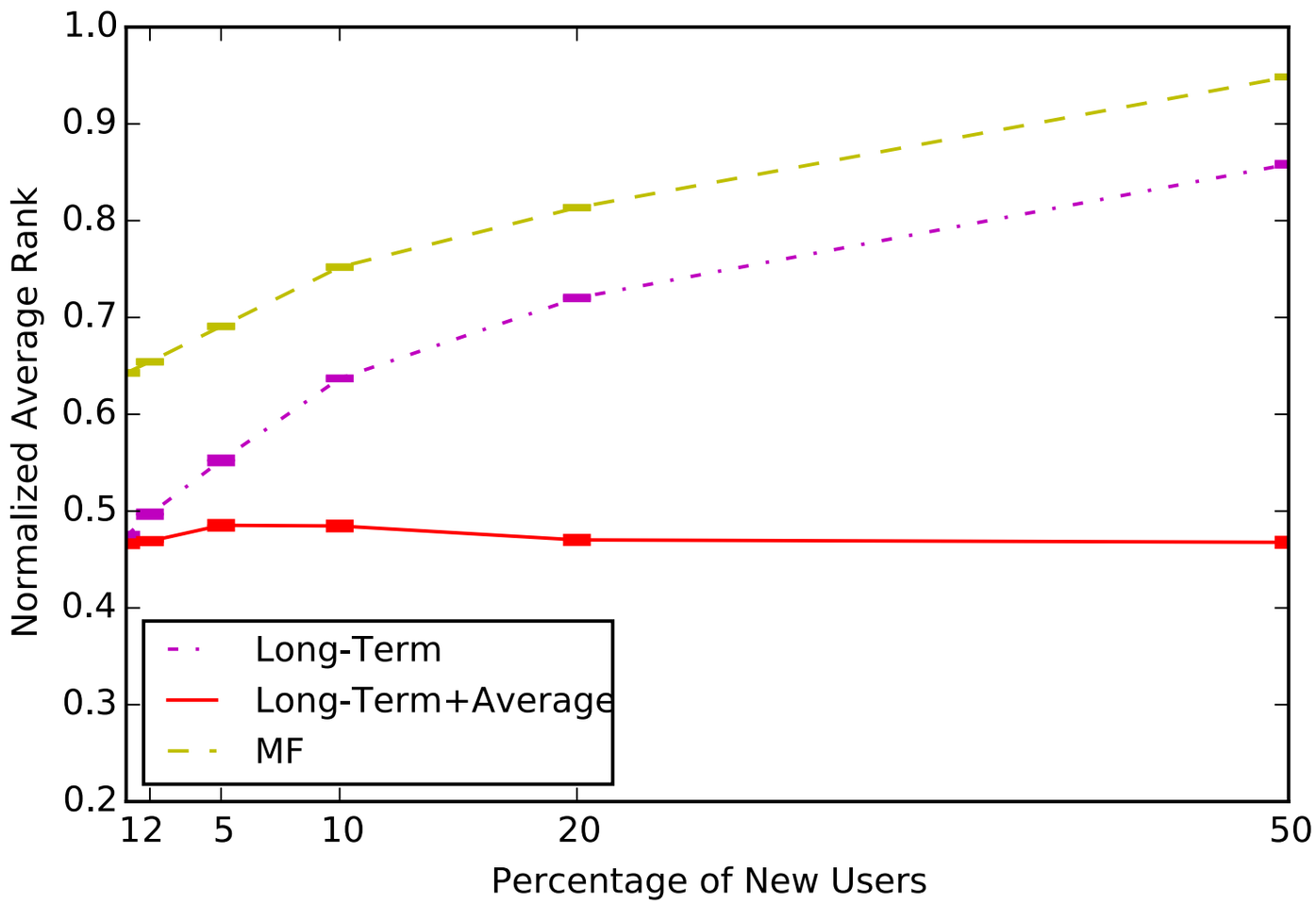
Short-Term Models

- The average term compensates for the new items



Long-Term Models

- The average term compensates for the new users



Conclusion

- The short- and long-term interests of users are combined in one model
- Free choice of *loss function* and *model complexity*
  - There is not one best model: the choice depends on the application

Questions?

Thanks for your attention

A Unified Contextual Bandit Framework for Long- and Short-Term Recommendations

Maryam Tavakol & Ulf Brefeld
{tavakol,brefeld}@leuphana.de

Source code available at https://github.com/marytavakol/Bandits