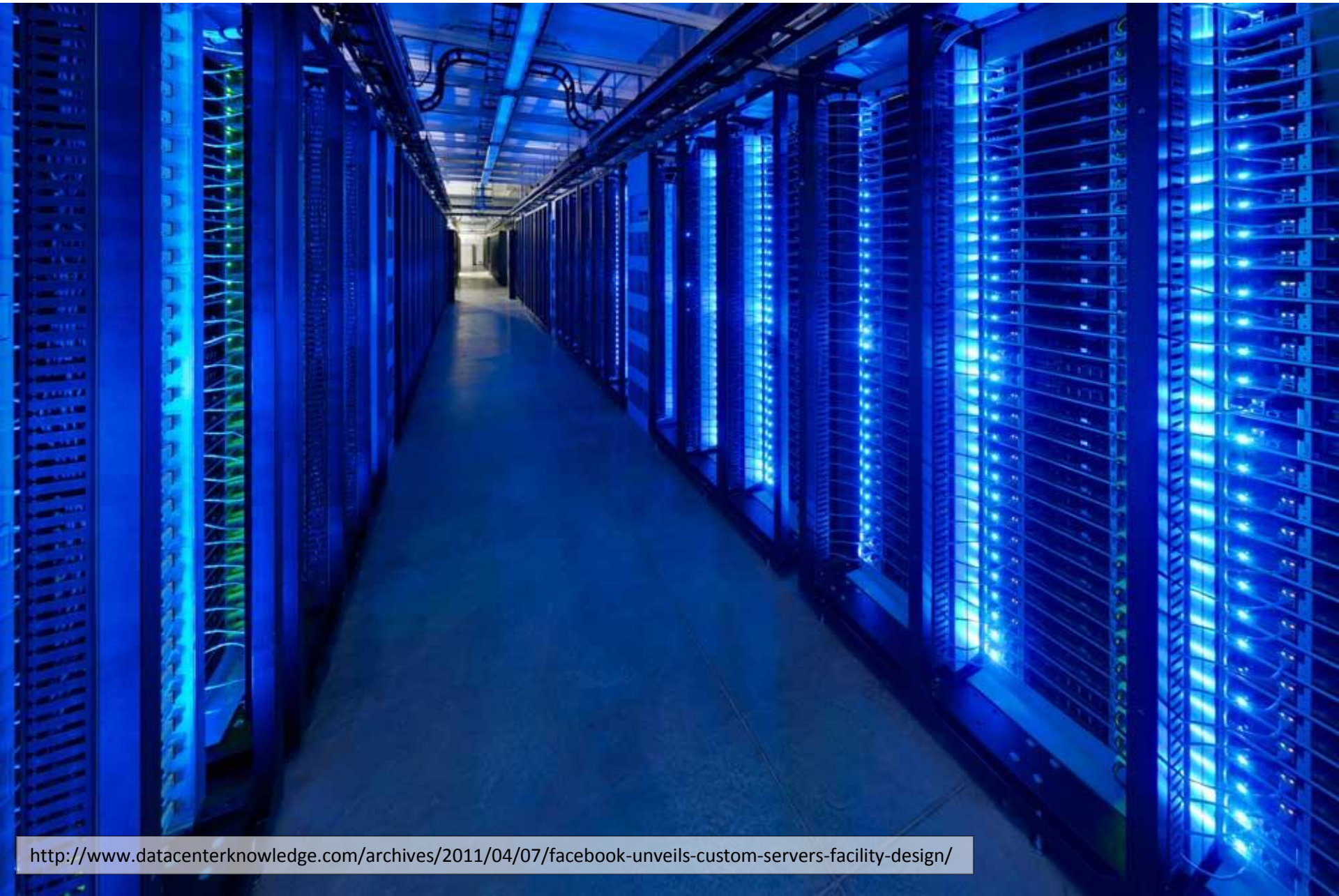
# Challenges in Online Learning to Rank for Information Retrieval (IR)

Katja Hofmann – Microsoft Research Cambridge*

*Work done with Shimon Whiteson and Maarten de Rijke at the Intelligent Systems Lab Amsterdam (ISLA), University of Amsterdam

# Labels for Large-Scale Learning



http://www.datacenterknowledge.com/archives/2011/04/07/facebook-unveils-custom-servers-facility-design/

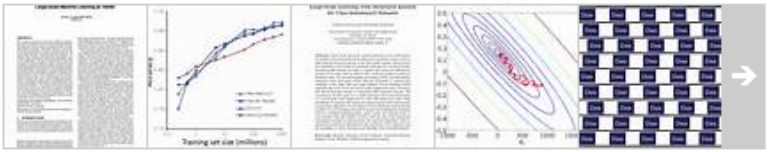
# Example: Web Search

large scale learning

ICML'08 Workshop PASCAL **Large Scale Learning** Challenge -- July 9, 2008
**largescale**.ml.tu-berlin.de/workshop
Pascal **Large Scale Learning** Challenge. ... 14:00 - 14:30: Han-Shen Huang and Chun-Nan Hsu - Triple Jump Linear SVM: abstract

Designing a **large scale learning** programme - 1
www.**learning**conversations.co.uk/main/...a-**large-scale-learning**...
When it comes to designing a **large-scale learning** programme, the processes you work through are probably not much different to building something just for your team.

Images of **large scale learning**
bing.com/images



**Large-Scale** Machine **Learning** - Computer Science & Engineering
www.cs.washington.edu/ai/lsml.html
**Large-Scale** Machine **Learning** Overview In many domains, data now arrives faster than we are able to **learn** from it. To avoid wasting this data, we must switch from the ...

**Large Scale Learning** - start [leon.bottou.org]
leon.bottou.org/research/**largescale**
The methods of conventional statistics were developed in times where both dataset collection and modeling were carried out with paper and pencil.

**Context:**
user, query

**Items:**
web documents

**Features:**
Match between query and document (e.g., TF-IDF), popularity (PageRank), match with user history, popularity in users' location, …

**Goal:**
Learn to rank documents to maximize utility to user

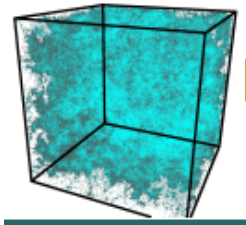
# Manual Annotations?

Query: **large scale learning**

ICML'08 Workshop PASCAL **Large Scale Learning** Challenge -- July 9, 2008

**largescale**.ml.tu-berlin.de/workshop

Pascal **Large Scale Learning** Challenge. ... 14:00 - 14:30: Han-Shen Huang and Chun-Nan Hsu - Triple Jump Linear SVM: abstract



**Pascal Large Scale Learning Challenge**

About | Instructions | Registration | Submission | Evaluation | Workshop | Summary | JMLR CfP

## Workshop

### ICML'08 Workshop PASCAL Large Scale Learning Challenge -- July 9, 2008

Topics: Large scale learning; Bounded-resource learning.

PASCAL
Pattern Analysis, Statistical Modelling and Computational Learning

**Motivation**

With the exceptional increase in computing power, storage capacity and network bandwidth of the past decades, ever growing datasets are collected in fields such as bioinformatics (Splice Sites, Gene Boundaries, etc), IT-security (Network traffic) or Text-Classification (Spam vs. Non-Spam), to name but a few. While the data size growth leaves computational methods as the only viable way of dealing with data, it poses new challenges to ML methods.

This workshop is concerned with the scalability and efficiency of existing ML approaches with respect to computational, memory or communication resources, e.g. resulting from a high algorithmic complexity, from the size or dimensionality of

# Labels for Large-Scale Learning

- Use observed labelled data as well as possible
  (e.g., semi-supervised learning)
- Crowdsourcing
- Learn directly from the environment
  (e.g., users of a system)

# Labels for Large-Scale Learning

- Use observed labelled data as well as possible
  (e.g., semi-supervised learning)
- Crowdsourcing
- **Learn directly from the environment**
  **(e.g., users of a system)**

# Learning from User Interactions: Challenges

Interpreting user interactions

Balancing exploration and exploitation

# Interpreting User Interactions

# Position Bias

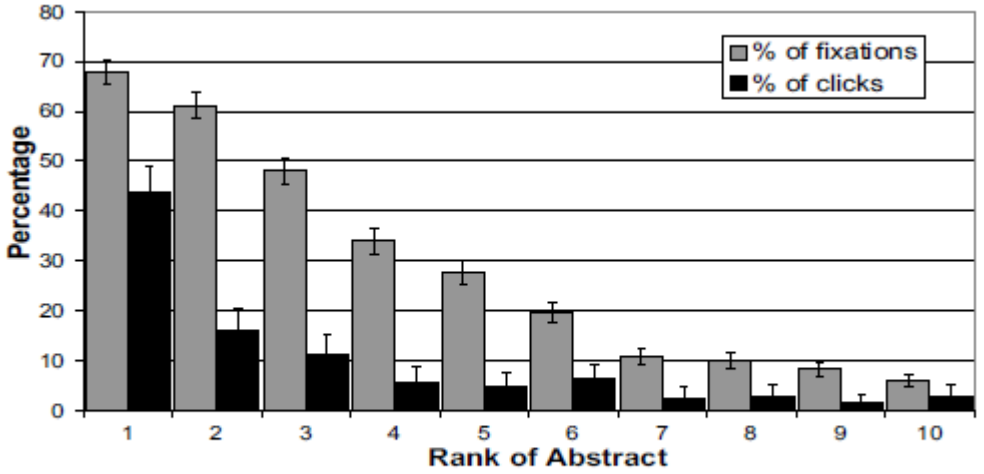



Figure 1: Percentage of times an abstract was viewed/clicked depending on the rank of the result.

T. Joachims, L. Granka, B. Pan, H. Hembrooke, G. Gay: *Accurately interpreting clickthrough data as implicit feedback.* SIGIR'05.

# Absolute Metrics

Absolute metrics typically used in A/B testing:

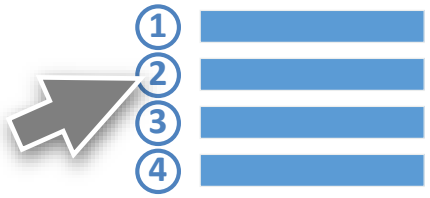
| | weak | | significant | |
|---|---|---|---|---|
| | ✓ | ↯ | ✓ | ↯ |
| Abandonment Rate | 4 | 2 | 2 | 0 |
| Reformulation Rate | 4 | 2 | 0 | 0 |
| Queries per Session | 3 | 3 | 0 | 0 |
| Clicks per Query | 4 | 2 | 2 | 0 |
| Clicks@1 | 4 | 2 | 4 | 0 |
| pSkip | 5 | 1 | 2 | 0 |
| Max Reciprocal Rank | 5 | 1 | 3 | 0 |
| Mean Reciprocal Rank | 5 | 1 | 2 | 0 |
| Time to First Click | 4 | 1 | 0 | 0 |
| Time to Last Click | 3 | 3 | 1 | 0 |

Number of correct (✓) and false (↯) preferences implied by absolute performance metrics in 6 large-scale experiments.
**Main finding:** None of the absolute metrics shows monotonic behaviour with changes in ranking quality.

O. Chapelle, T. Joachims, F. Radlinski, and Y. Yue: *Large-scale validation and analysis of interleaved search evaluation*. TOIS, Vol. 30, 2012.

# Interpreting clicks as **pairwise** feedback

Example: search engine returned 4 documents

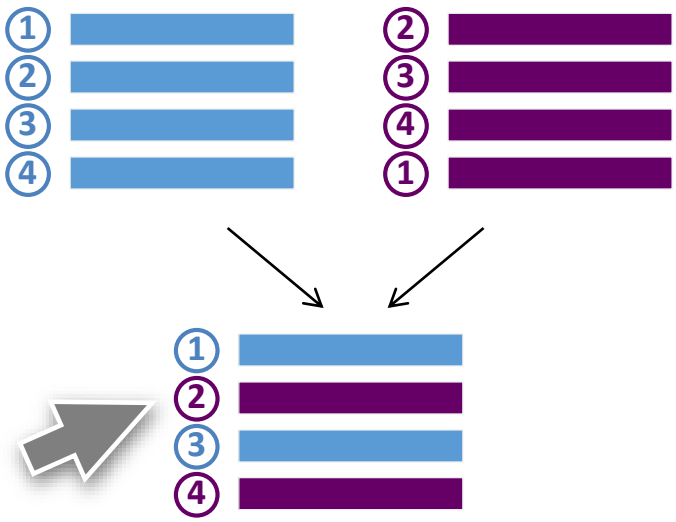



User skips document 1, clicks document 2
=> Interpret as a preference of 2 over 1.

T. Joachims, L. Granka, B. Pan, H. Hembrooke, G. Gay: *Accurately interpreting clickthrough data as implicit feedback.* SIGIR'05.

# Interpreting clicks as **listwise** feedback

**Interleaved comparison methods** compare result lists:




- Goal: Compare two result lists using click data
- Procedure:
  1) Generate interleaved result list (randomize per pair of ranks)
  2) Observe user clicks
  3) Credit clicks to original rankers to infer outcome $o \in \{-1, 0, +1\}$

F. Radlinski, M. Kurup, and T. Joachims: *How does clickthrough data reflect retrieval quality?*. CIKM'08.

# Summary: Interpreting interactions as feedback for learning

For web search and query suggestion ranking:

- Explicit labels may be biased due to how the labelling task is set up
- User interactions are difficult to interpret due to position (and other) bias

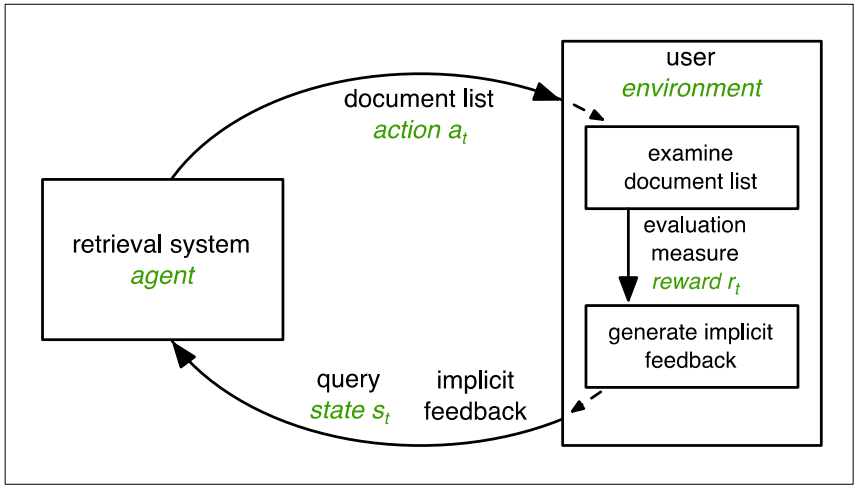
Promising direction: interpret user interactions as relative feedback (pairwise or listwise).

# Balancing exploration and exploitation

K. Hofmann, S. Whiteson, M. de Rijke*: Balancing Exploration and Exploitation in Listwise and Pairwise Online Learning to Rank for Information Retrieval*. Information Retrieval, Vol 16, 2012.

K. Hofmann, S. Whiteson, M. de Rijke: *Balancing Exploration and Exploitation in Learning to Rank Online*. ECIR'11.

# Problem Formulation




The IR problem modeled as a contextual bandit problem with RL terminology in *green* and IR terminology in black.

**Reinforcement learning (RL) Approach**

**L**earn by trying out actions (document lists), and observing implicit feedback

- **Formulation:** contextual bandit problem
  - context = features for query – document pairs: $\mathbf{X} = \{\mathbf{x}_1, ...\mathbf{x}_D\}$
  - queries are independent
- **Goal:** present result lists $\mathbf{l}_t$ that maximize discounted cumulative reward:

$$C = \sum_{t=0}^{\infty} \gamma^{t-1} r_t(\mathbf{l}_t)$$

$\gamma$ - discount factor

# The Online Learning to Rank Challenge

Obtain feedback that is useful for learning

Keep users happy (present high-quality results) while learning

➔ **Exploration**

➔ **Exploitation**

Previous learning to rank approaches for IR are either purely exploratory or purely exploitative.

User behavior in IR is difficult to interpret due to (rank) bias and noise.

# Question

- Can online learning to rank for IR be improved by balancing exploration and exploitation?

# Approach

- Extend two online learning methods (**pairwise** and **listwise**) to allow balancing exploration and exploitation
- Study performance under different settings + assumptions

# Pairwise Learning to Rank for IR

**Input:**

feature vectors constructed from document pairs

$(\mathbf{x}(q, d_i), \mathbf{x}(q, d_j)) \in \mathbb{R}^n \times \mathbb{R}^n$

**Output:**

$y \in \{-1, +1\}$ (incorrect / correct order)

**Goal:**

Learn mapping using any supervised learning method

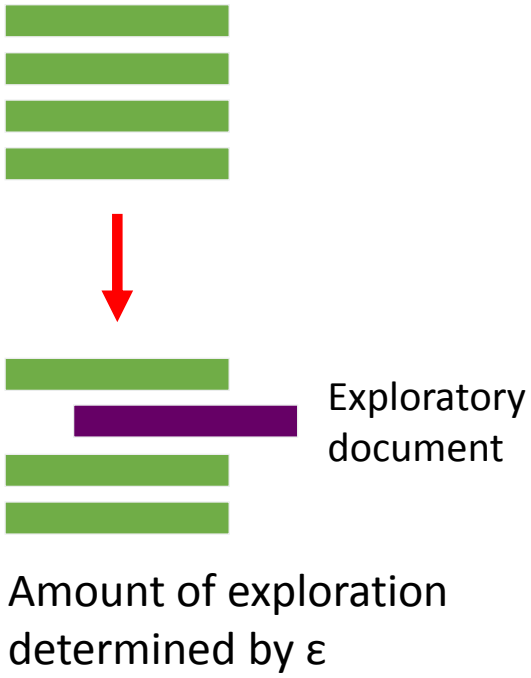
here: stochastic gradient descent, with update rule

$$\textbf{if } y\mathbf{w}_{t-1}^T(\mathbf{x}_1 - \mathbf{x}_2) < 1.0 \text{ and } y \neq 0.0 \textbf{ then}$$
$$\mathbf{w}_t = \mathbf{w}_{t-1} + \eta y(\mathbf{x}_1 - \mathbf{x}_2) - \eta\lambda\mathbf{w}_{t-1}$$

# Balancing Exploration and Exploitation in Pairwise Learning

**Baseline [1]:**
Present ranking based on current weight vector (**purely exploitative**)

**Idea:**
Adapt **ε-greedy**.
At each rank, select the next exploitative document with p = 1- ε; select a randomly sampled (exploratory) candidate document with p = ε




[1] D. Sculley: *Large scale learning to rank*. NIPS'09 Workshop on Advances in Ranking.

# Listwise Learning to Rank for IR

**Input:**

Feature vectors for all candidate documents

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_1, ...\mathbf{x}_D\}$$

**Output:**

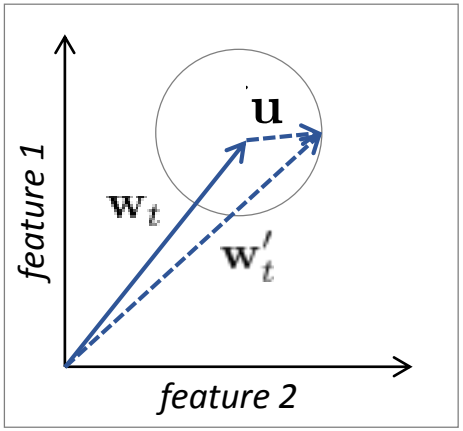
Complete ranking of the candidate documents

(by a score $S = \mathbf{wx}(q, d)$)

**Learning approach:**

Has to work with listwise relative feedback

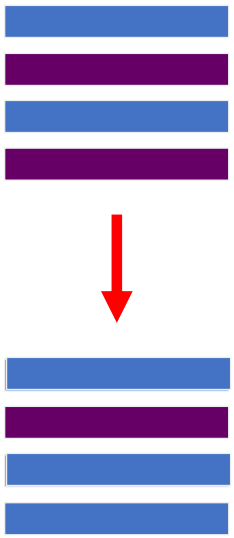
here: stochastic gradient descent ("Dueling Bandits")



# Balancing Exploration and Exploitation in Listwise Learning

**Baseline [1]:**
Compare $\mathbf{w}_t$ and $\mathbf{w}'_t$ using interleaved comparisons (**purely exploratory**)

**Idea:**
Allow "statistical" interleaving. Introduce a parameter (**k**) that determines **ratio of exploitative and exploratory documents** (compensate for bias due to document ratios after observing clicks)



Original interleaving

Statistical interleaving

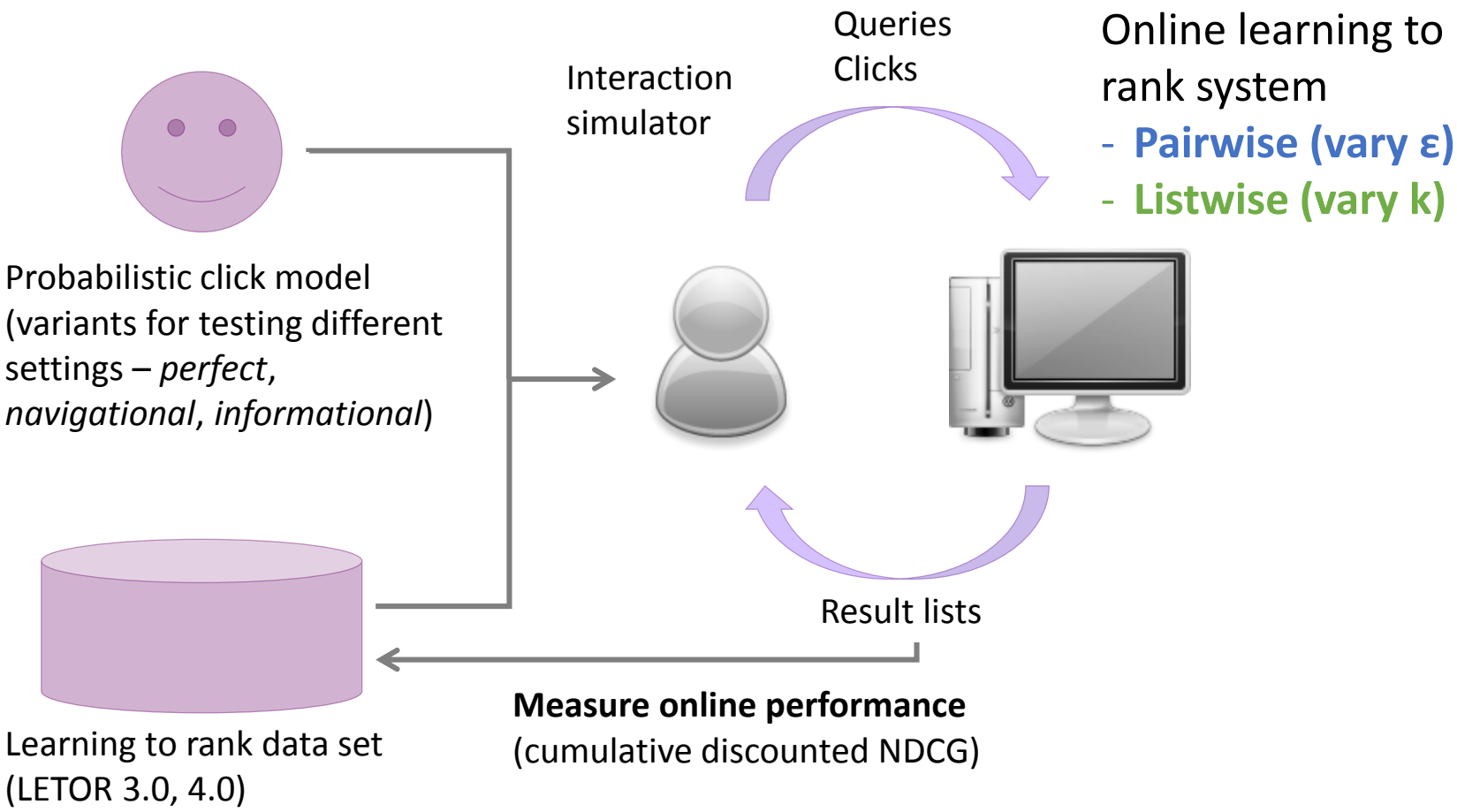
Ratio of exploitative and exploratory documents determined by k

[1] Y. Yue and T. Joachims: *Interactively optimizing information retrieval systems as a dueling bandits problem.* ICML'09.

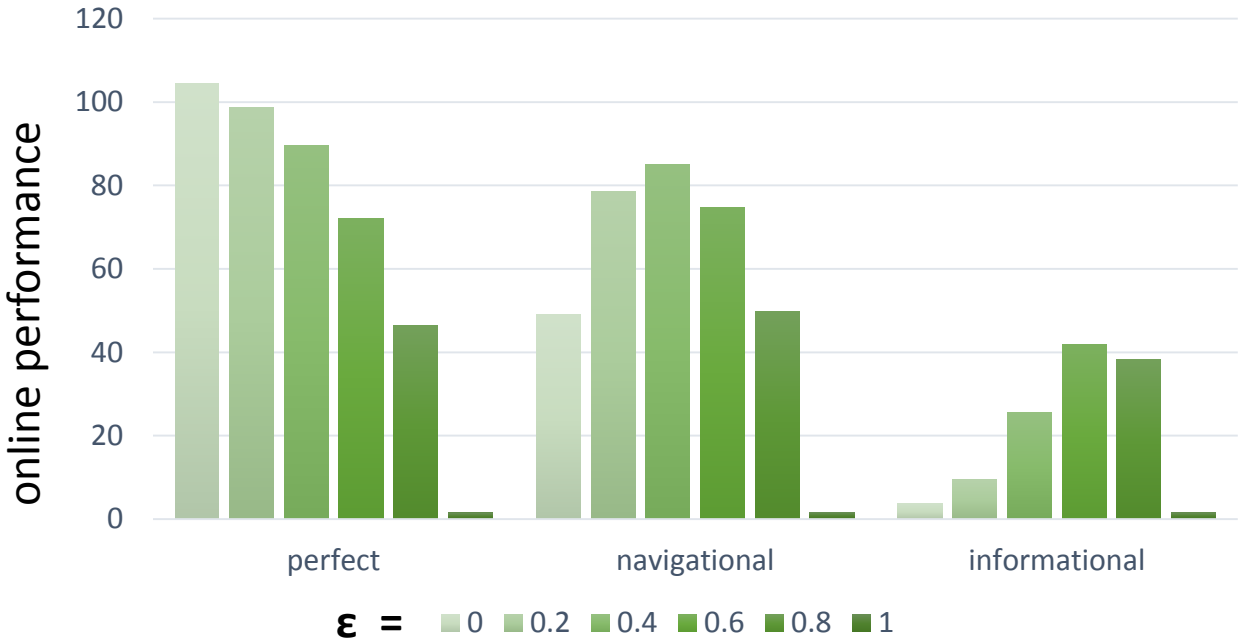
# 2 Approaches – Summary

- **Pairwise** – learns preferences between documents
  - Exploit by presenting the current “best” ranker
  - Explore by injecting random documents
- **Listwise** – learns preferences for complete rankings
  - Explore by interleaving exploratory and exploitative list in equal parts
  - Exploit by showing more exploitative documents

# Experiments



# Results: **Online** Performance
# Pairwise Approach

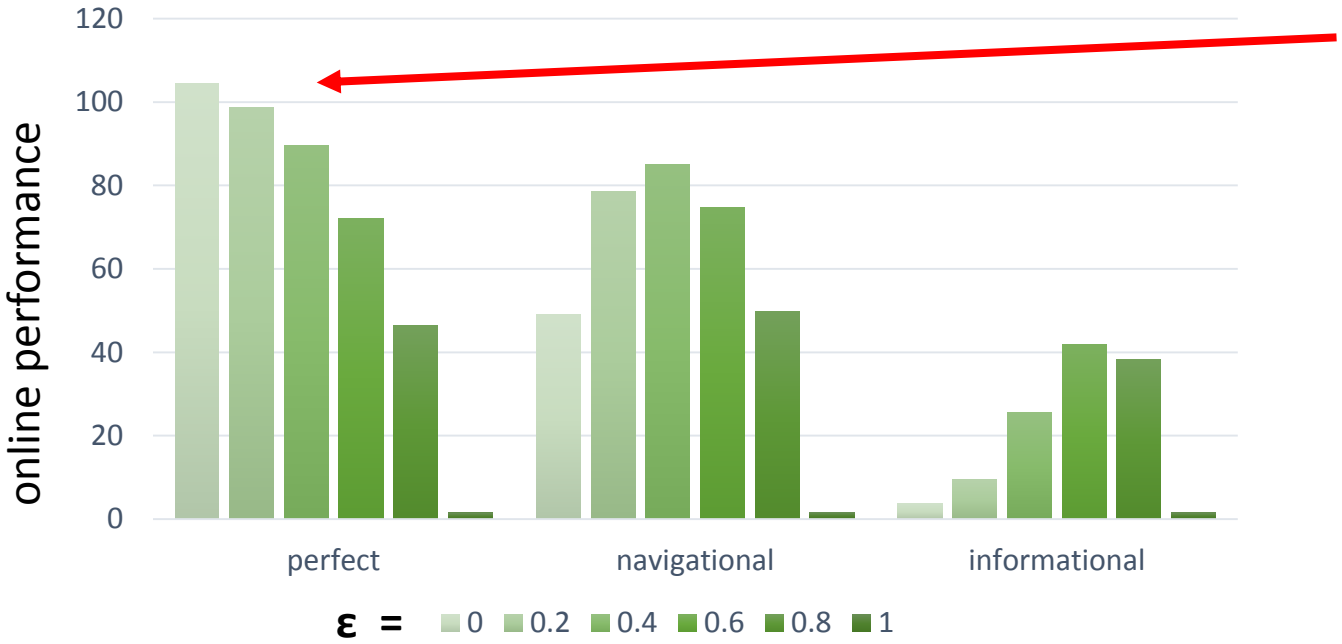



pure exploitation

pure exploration

(data set: NP2003)

# Results: **Online** Performance
# Pairwise Approach



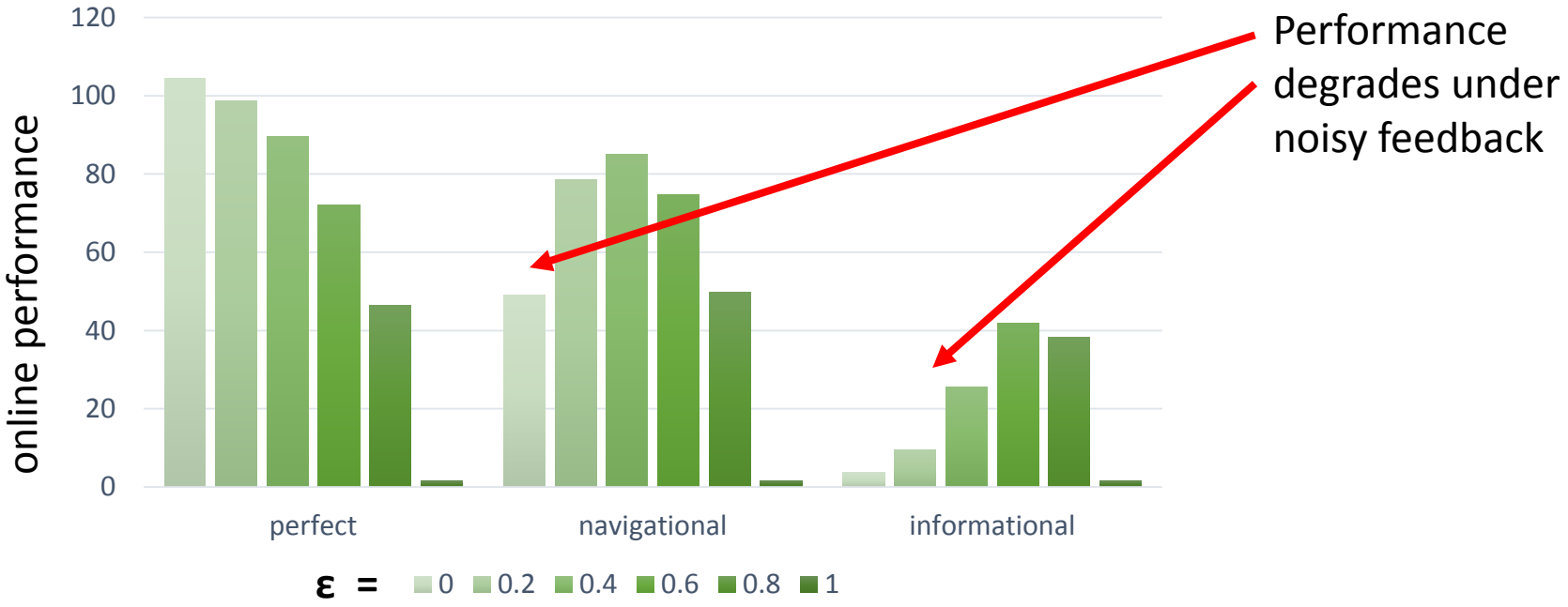

High online performance in exploitative setting when feedback is very reliable

pure exploitation

pure exploration

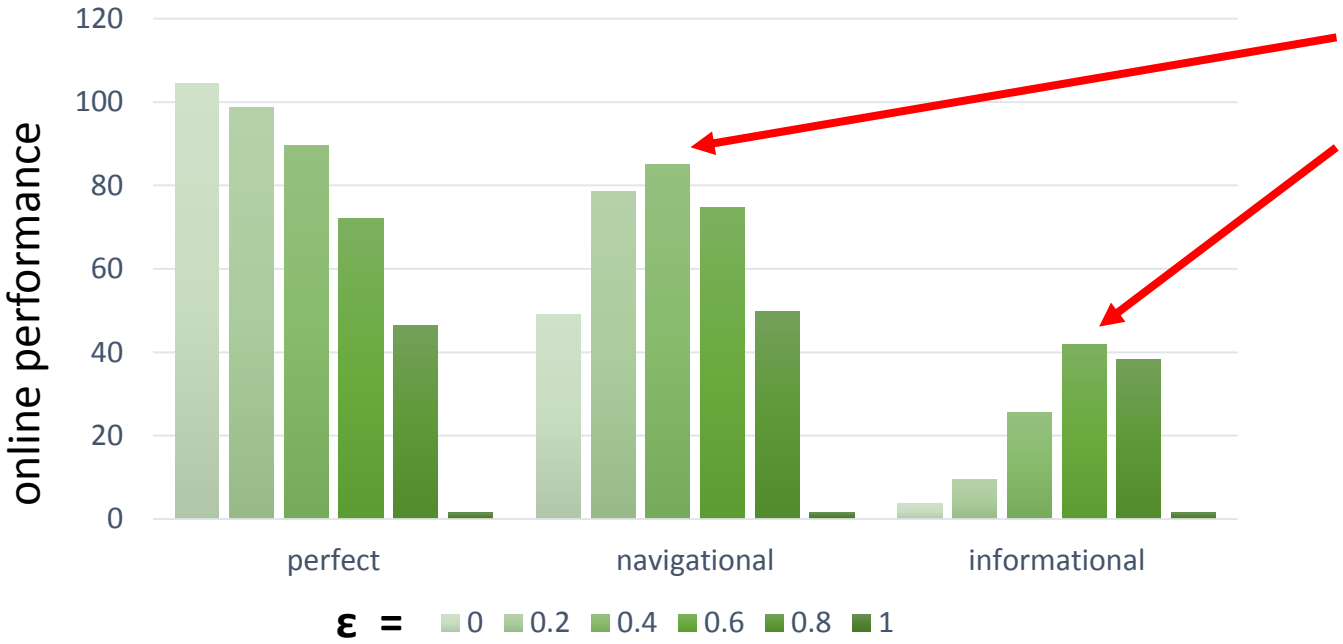
(data set: NP2003)

# Results: **Online** Performance
# Pairwise Approach




(data set: NP2003)

# Results: **Online** Performance
# Pairwise Approach



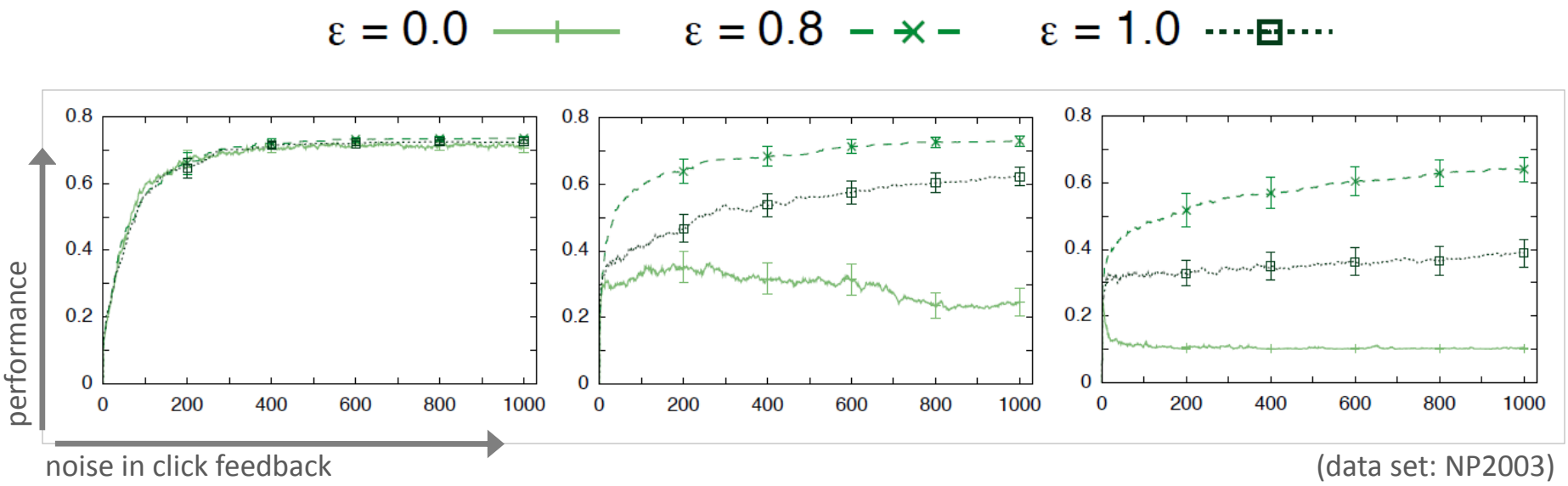

Increasing exploration can compensate for feedback noise to some degree

pure exploitation

pure exploration

(data set: NP2003)

# Results: **Offline** Performance
# Pairwise Approach

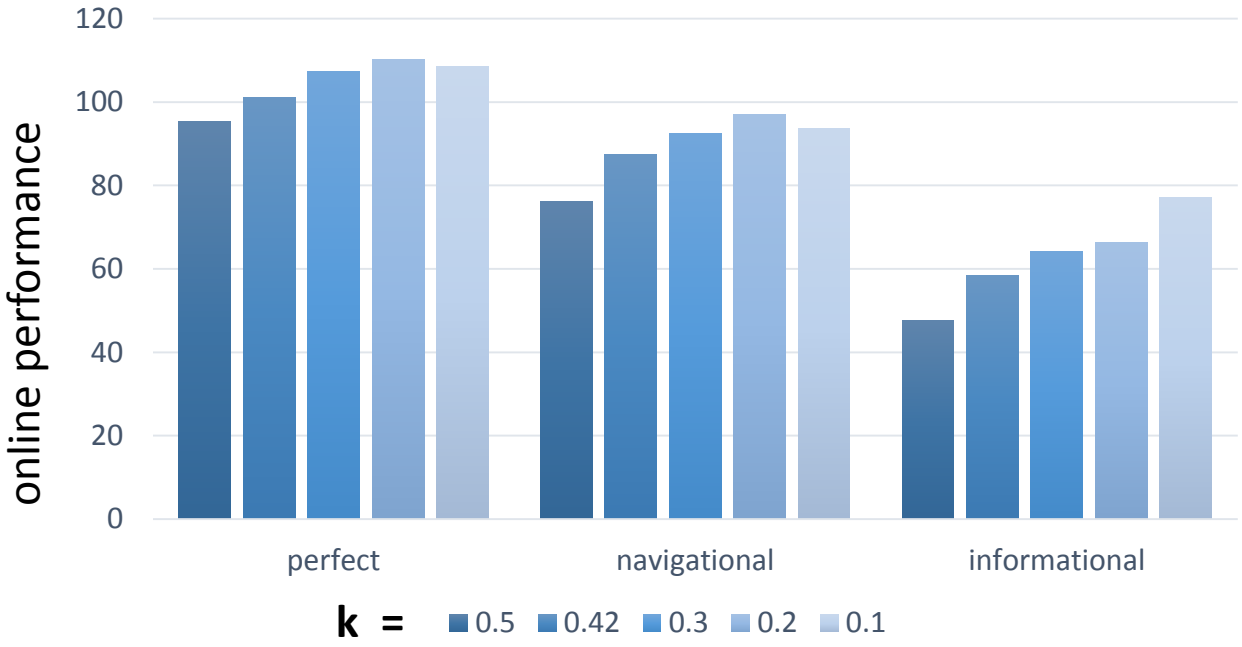



(data set: NP2003)

Very effective learning under reliable feedback (irrespective of exploration rate)

High level of exploration needed to counter noisy feedback

# Results: **Online** Performance
# Listwise Approach

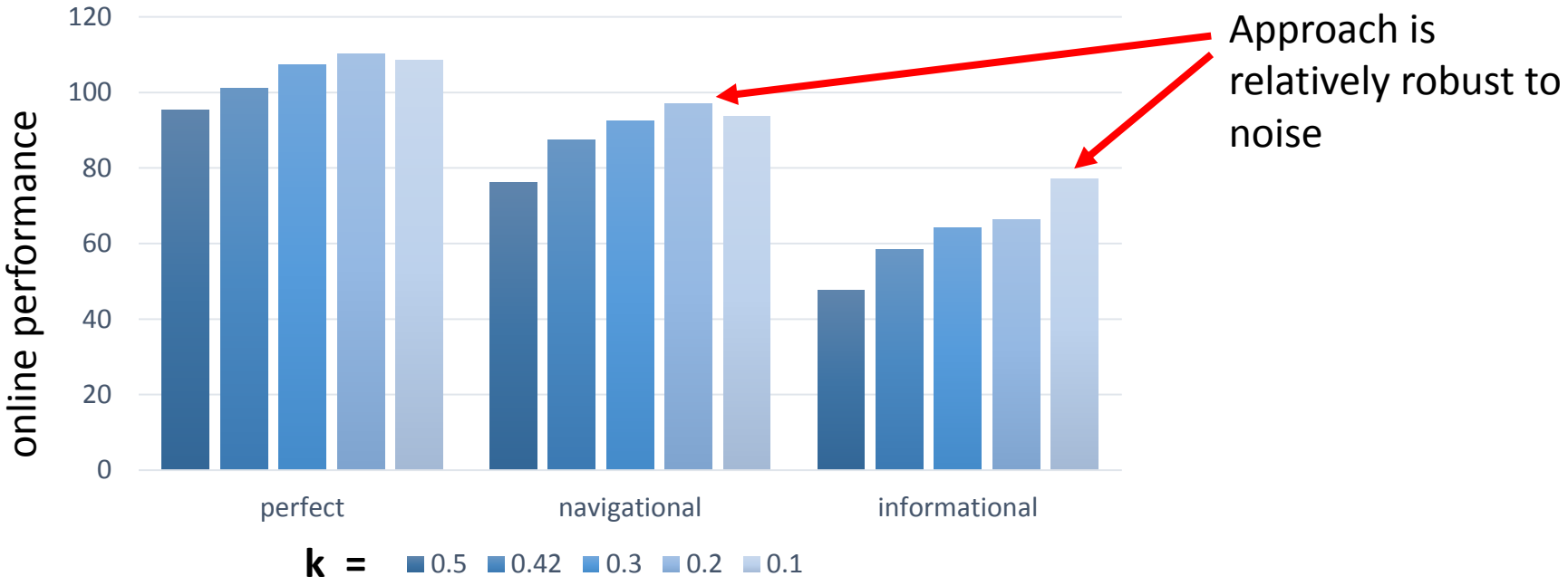



pure exploration

pure exploitation

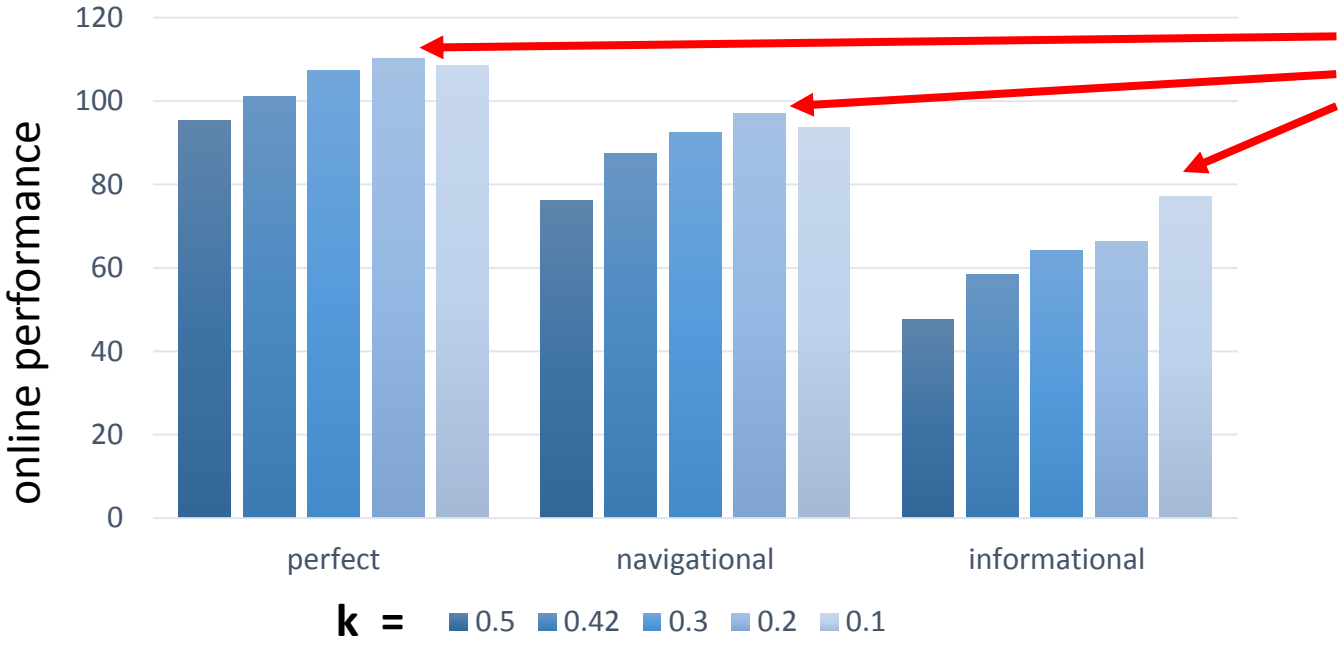
(data set: NP2003)

# Results: **Online** Performance
# Listwise Approach




(data set: NP2003)

# Results: **Online** Performance
# Listwise Approach



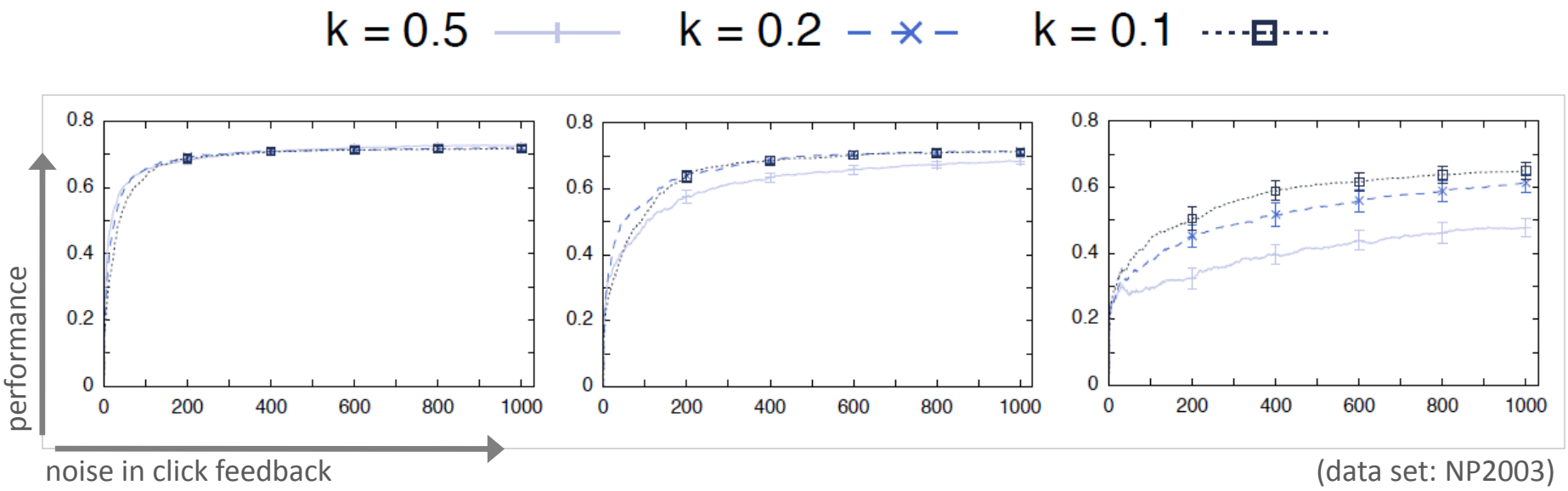

Highest performance achieved in more exploitative settings (irrespective of noise)

pure exploration

pure exploitation

(data set: NP2003)

# Results: **Offline** Performance
# Listwise Approach




(data set: NP2003)

Baseline approach over-explores – need to decrease exploration rate for optimal performance

Much more robust to noise than pairwise approach

# Summary

**Learning from user interactions**

Most promising: interpretations as relative feedback for learning

**Balancing exploration and exploitation:**
improves online performance

Optimal balance depends on approach, level of noise in click feedback

Code: https://bitbucket.org/ilps/lerot.git
Documentation: A. Schuth, K. Hofmann, S. Whiteson and M. de Rijke: *Lerot: An online learning to rank framework*. Living Lab'13.

# Related / Ongoing Work

## Probabilistic interleave

Infers interleaved comparison outcomes based on graphical model

K. Hofmann, S. Whiteson, M. de Rijke: *A Probabilistic Method for Inferring Preferences from Clicks*. CIKM'11.

Allows data reuse

K. Hofmann, S. Whiteson, M. de Rijke: *Estimating Interleaved Comparison Outcomes from Historical Click Data*. CIKM'12.

## Learning with probabilistic interleave

Data reuse can reduce required exploration / substantially speed learning

K. Hofmann, A. Schuth, S. Whiteson, M. de Rijke: *Reusing Historical Interaction Data for Faster Online Learning to Rank for IR*. WSDM'13.

# Outlook: Smart Exploration


**So far:** Balancing exploration and exploitation improves online performance, but exploration is random

**Idea:** Explore several promising areas of the solution space in parallel, utilize historical data to zoom in on promising areas

**Key challenges:** How to compare large sets of rankers as efficiently as possible? How to model solution spaces for ranking?