UNIVERSITY OF AMSTERDAM

Optimizing Base Rankers Using Clicks

A Case Study using BM25

Anne Schuth, Floor Sietsma, Shimon Whiteson, and Maarten de Rijke

Outline

- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

Typical Learning to Rank Scenario

Typical Learning to Rank Scenario

- Select a set of representative queries

Typical Learning to Rank Scenario

- Select a set of representative queries
- Assess top documents for each of these queries

Typical Learning to Rank Scenario

- Select a set of representative queries
- Assess top documents for each of these queries
- Select many (dozens, hundreds) of features:
  - Base rankers: PageRank, BM25, Language Models, …
  - Each base ranker applied to title of document, body of document, url, …

Typical Learning to Rank Scenario

- Select a set of representative queries
- Assess top documents for each of these queries
- Select many (dozens, hundreds) of features:
  - Base rankers: PageRank, BM25, Language Models, …
  - Each base ranker applied to title of document, body of document, url, …
- Compute feature values for all query-document pairs

Typical Learning to Rank Scenario

- Select a set of representative queries
- Assess top documents for each of these queries
- Select many (dozens, hundreds) of features:
  - Base rankers: PageRank, BM25, Language Models, …
  - Each base ranker applied to title of document, body of document, url, …
- Compute feature values for all query-document pairs
- Train a combination of features that maximizes some metric
  - Combinations: Linear, neural net, …
  - Metrics: nDCG, MAP, …

Typical Learning to Rank Scenario

- Select a set of representative queries
- Assess top documents for each of these queries
- Select many (dozens, hundreds) of features:
  - Base rankers: PageRank, BM25, Language Models, …
  - Each base ranker applied to title of document, body of document, url, …
- Compute feature values for all query-document pairs
- Train a combination of features that maximizes some metric
  - Combinations: Linear, neural net, …
  - Metrics: nDCG, MAP, …
- Apply the learned model in the wild

Two issues

Two issues

1. Assessments

Two issues

1. Assessments
   - expensive (laborious)

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users
   - …

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users
   - …
2. Choosing base rankers parameters

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users
   - …
2. Choosing base rankers parameters
   - Don't choose (choose many)

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users
   - …
2. Choosing base rankers parameters
   - Don’t choose (choose many)
   - From literature

Two issues

1. Assessments
   - expensive (laborious)
   - incomplete (selection of queries, selection of documents)
   - quickly outdated
   - come from assessors instead of users
   - …
2. Choosing base rankers parameters
   - Don’t choose (choose many)
   - From literature
   - Using the learning to rank procedure

Solutions to these issues

Solutions to these issues

1. Assessments

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date
   - But: biased

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date
   - But: biased
     - Online learning (with exploration) to rank (instead of supervised)

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date
   - But: biased
     - Online learning (with exploration) to rank (instead of supervised)
2. Choosing base rankers parameters

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date
   - But: biased
     - Online learning (with exploration) to rank (instead of supervised)
2. Choosing base rankers parameters
   - Using the online learning to rank procedure

Solutions to these issues

1. Assessments
   - Use implicit user feedback (clicks) instead
     - Widely available (in industry)
     - Natural byproduct
     - Always up to date
   - But: biased
     - Online learning (with exploration) to rank (instead of supervised)
2. Choosing base rankers parameters
   - Using the online learning to rank procedure
   - But base rankers are not always linear

Outline

- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

A Case Study using BM25

A Case Study using BM25

- RQ1: How good are the manually tuned parameter values of BM25 that are currently used?

A Case Study using BM25

- RQ1: How good are the manually tuned parameter values of BM25 that are currently used?
- RQ2: Are they optimal for all data sets on average? Are they optimal for individual data sets?

A Case Study using BM25

- RQ1: How good are the manually tuned parameter values of BM25 that are currently used?
- RQ2: Are they optimal for all data sets on average? Are they optimal for individual data sets?
- RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

A Case Study using BM25

- RQ1: How good are the manually tuned parameter values of BM25 that are currently used?
- RQ2: Are they optimal for all data sets on average? Are they optimal for individual data sets?
- RQ3: Is it possible to learn good values of the BM25 parameters from clicks?
- RQ4: Can we approximate or even improve the performance of BM25 achieved with manually tuned parameters?

Outline

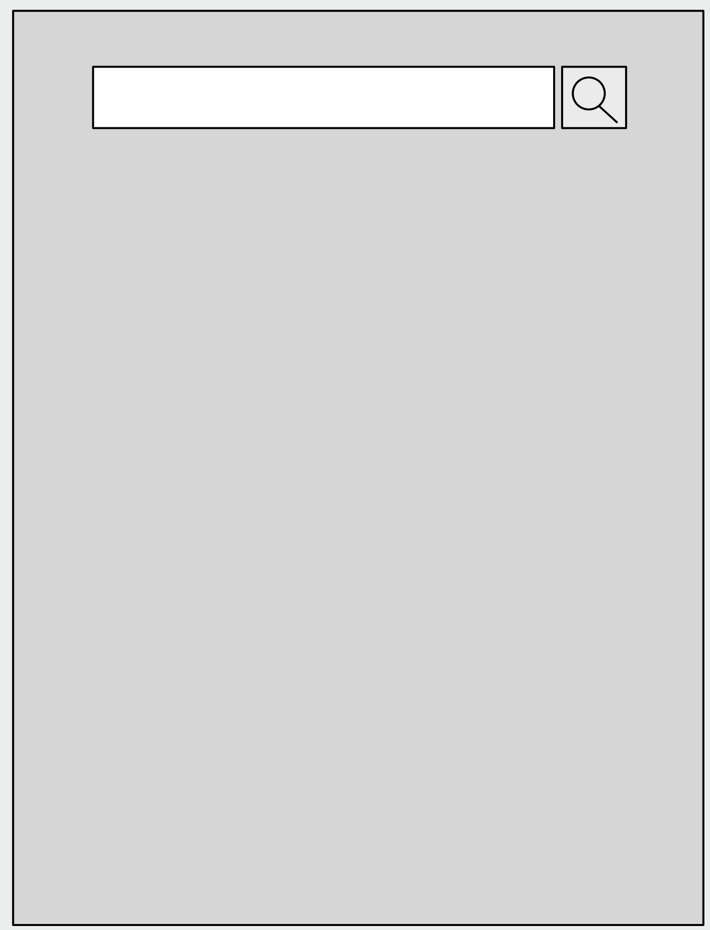
- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

A Case Study using BM25

$$BM25(q,d) = \sum_{q_i : tf(q_i,d)>0} \frac{idf(q_i) \cdot tf(q_i,d) \cdot (k_1+1)}{tf(q_i,d) + k_1 \cdot (1 - b + b \cdot \frac{|d|}{avgdl})}$$

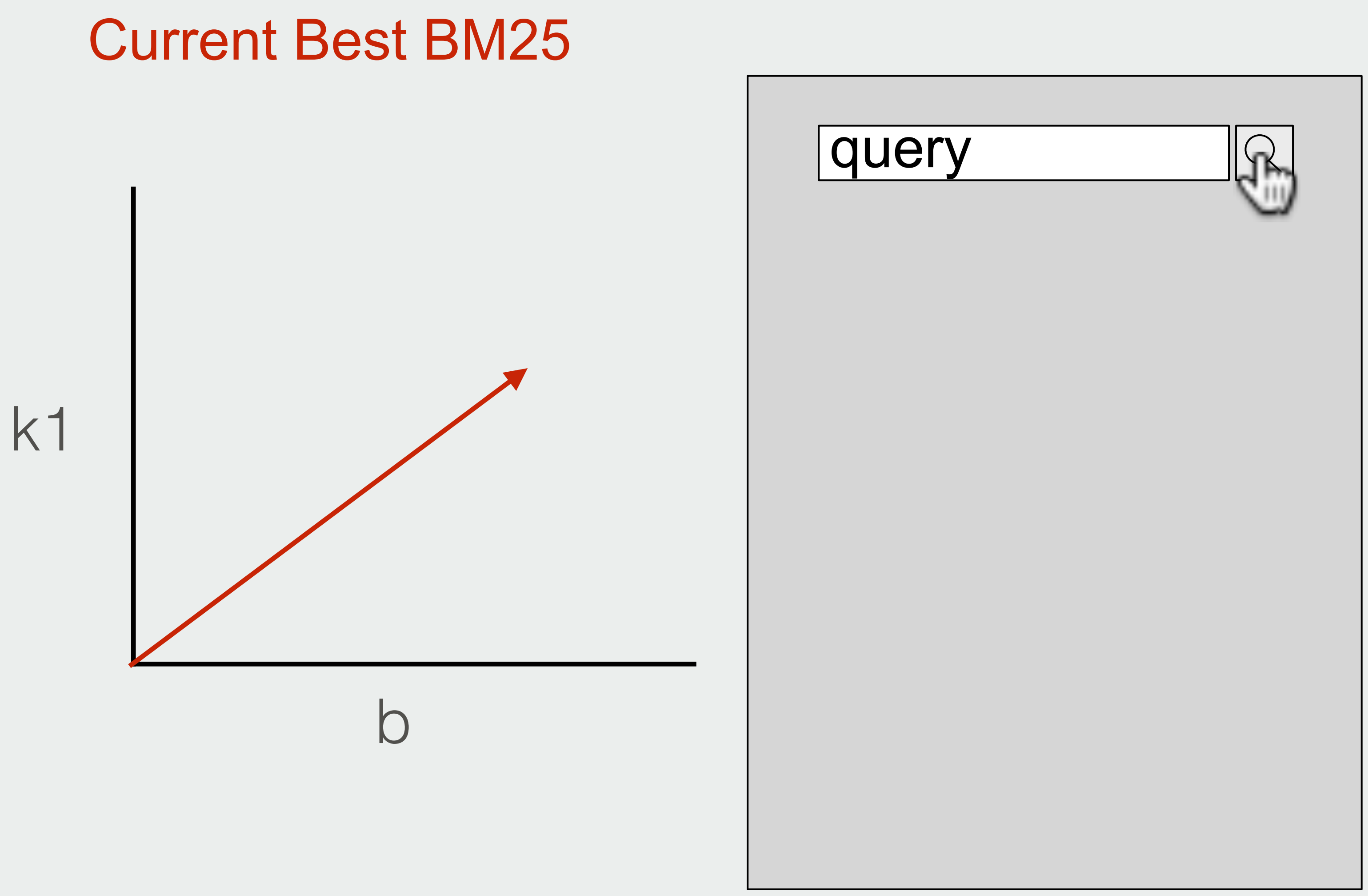
- We optimize 2 parameters: k_1 and b
- Typical magnitudes:
  - b between 0.45 and 0.9
  - k_1 between 2 and 25

Dueling Bandit Gradient Descent



[Yue et al, 2009; Hofmann et al., 2011]

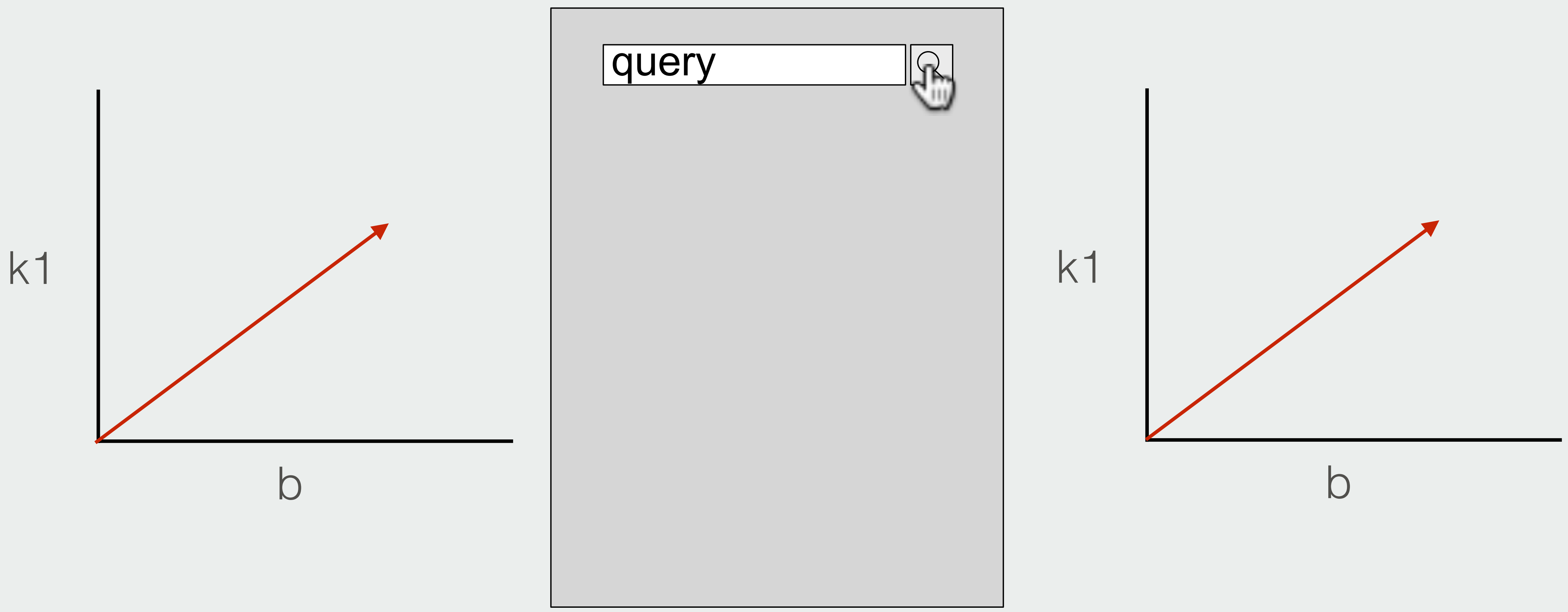
Dueling Bandit Gradient Descent



[Yue et al, 2009; Hofmann et al., 2011]

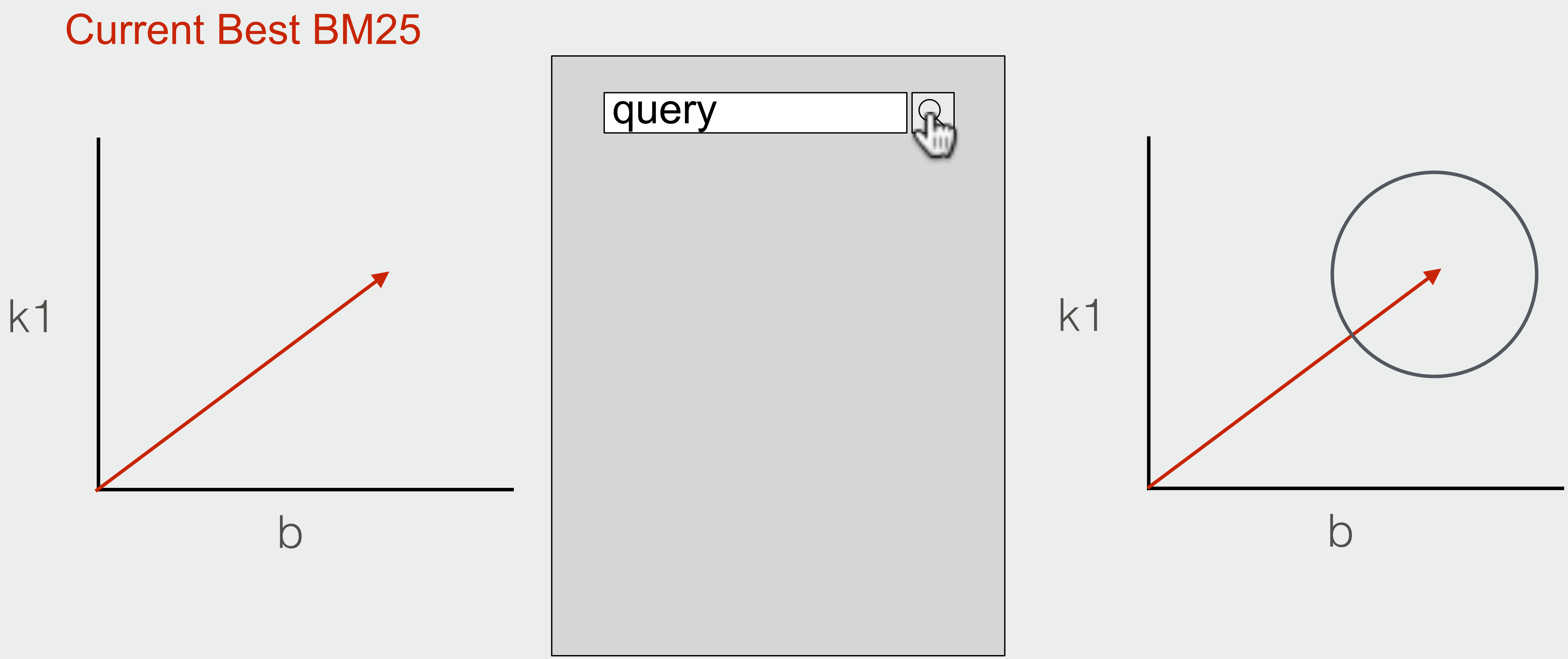
Dueling Bandit Gradient Descent

Current Best BM25



[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent

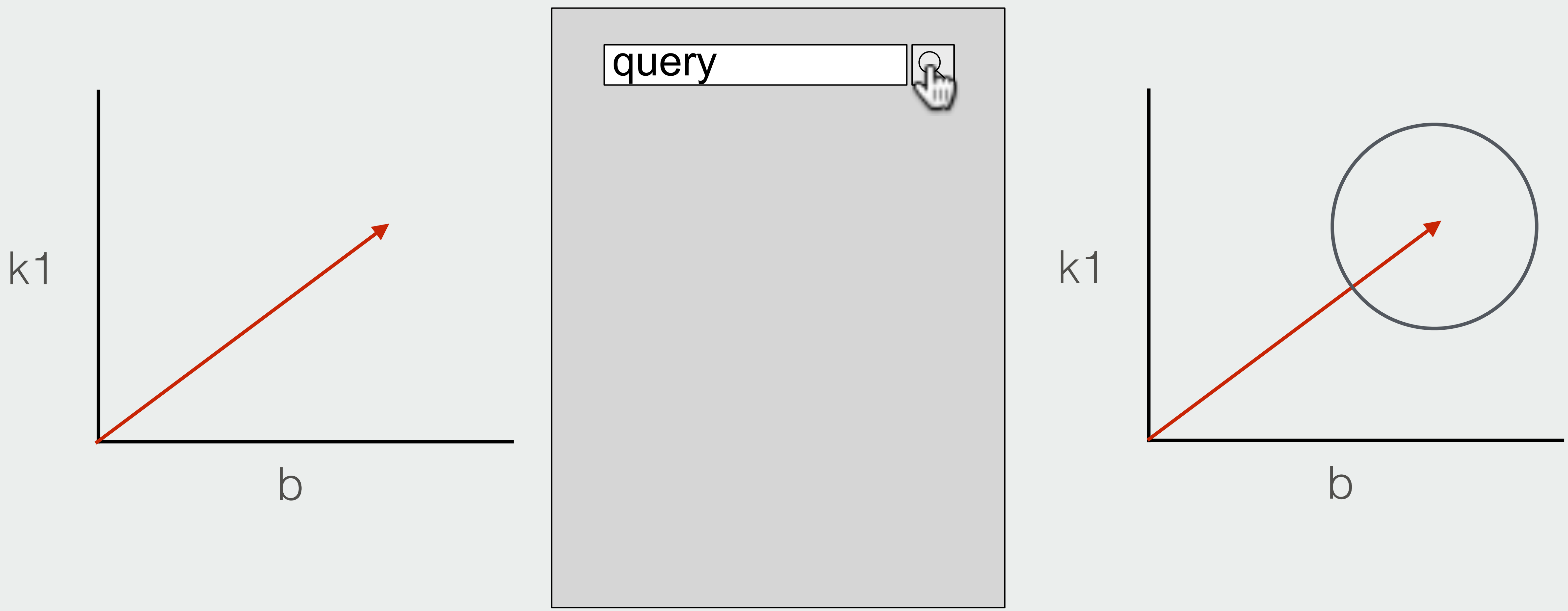


[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Des

k1 step size is larger then
step size of b

Current Best BM25

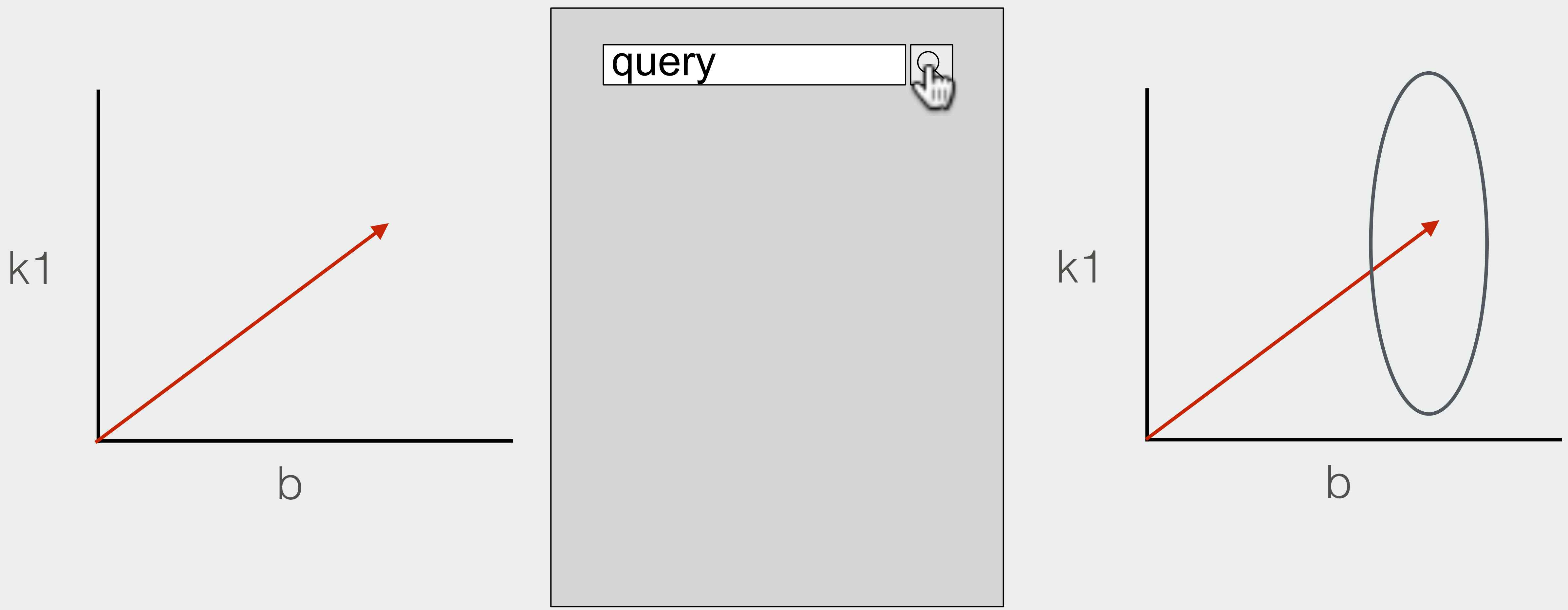


[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Des

k1 step size is larger then step size of b

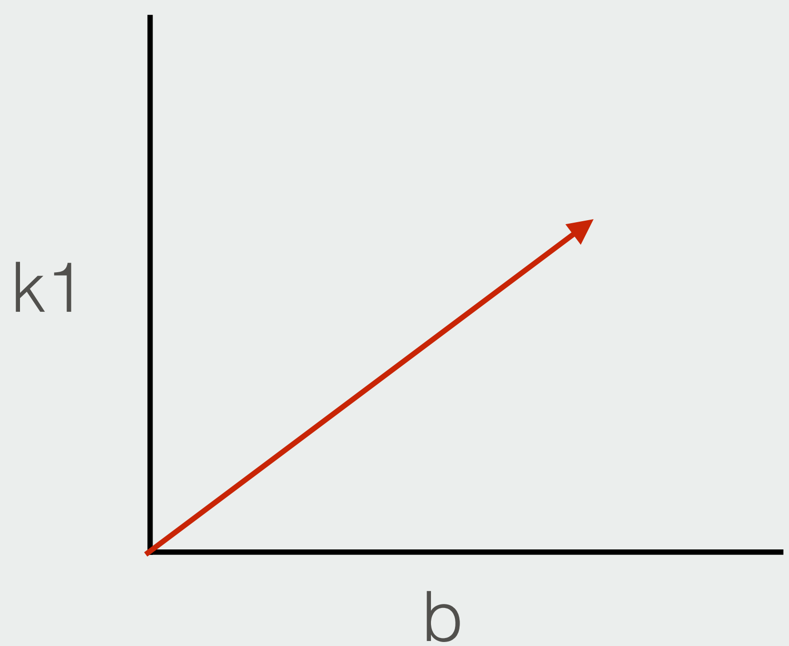
Current Best BM25

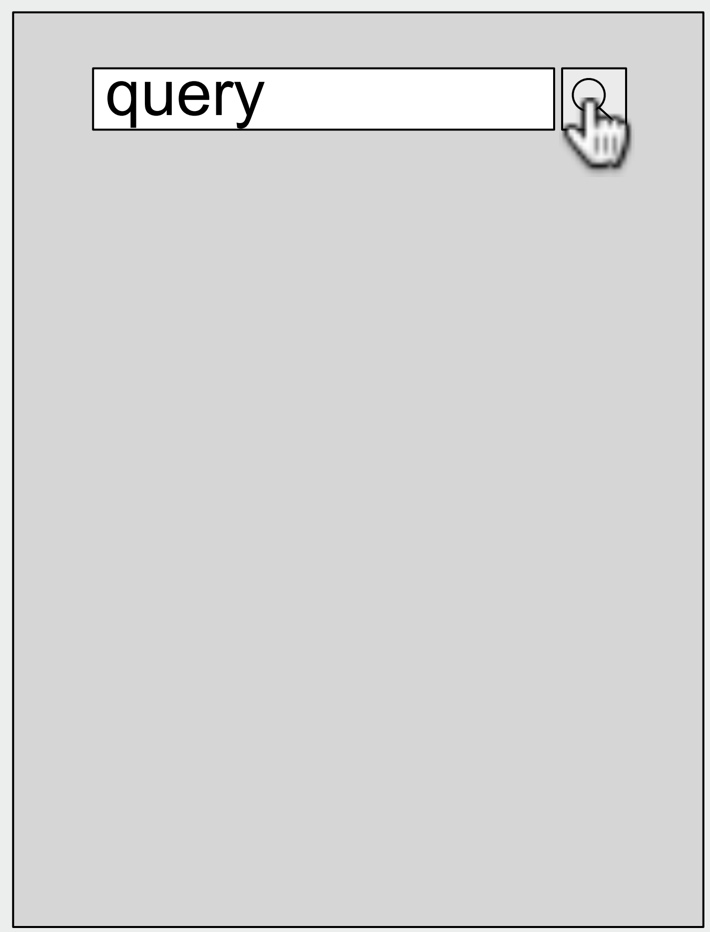


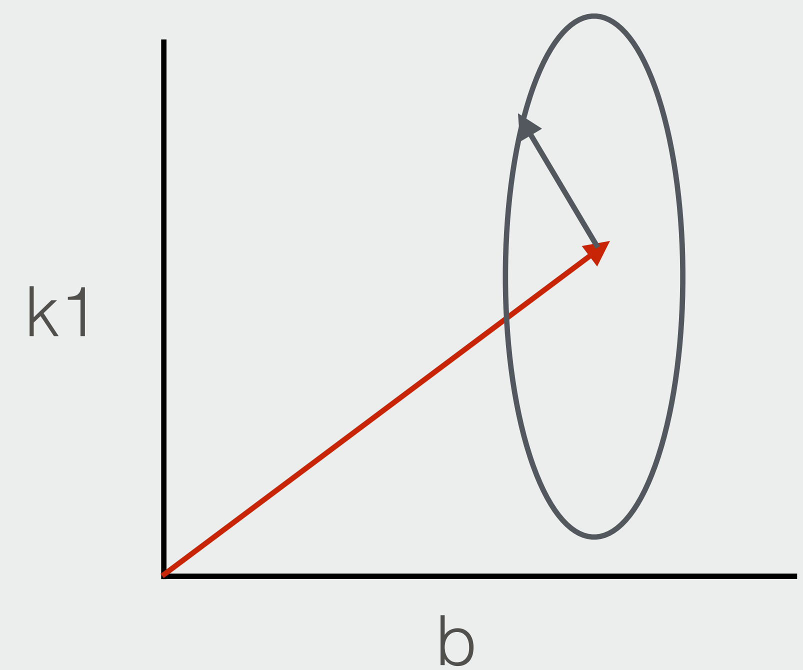
[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent

Current Best BM25



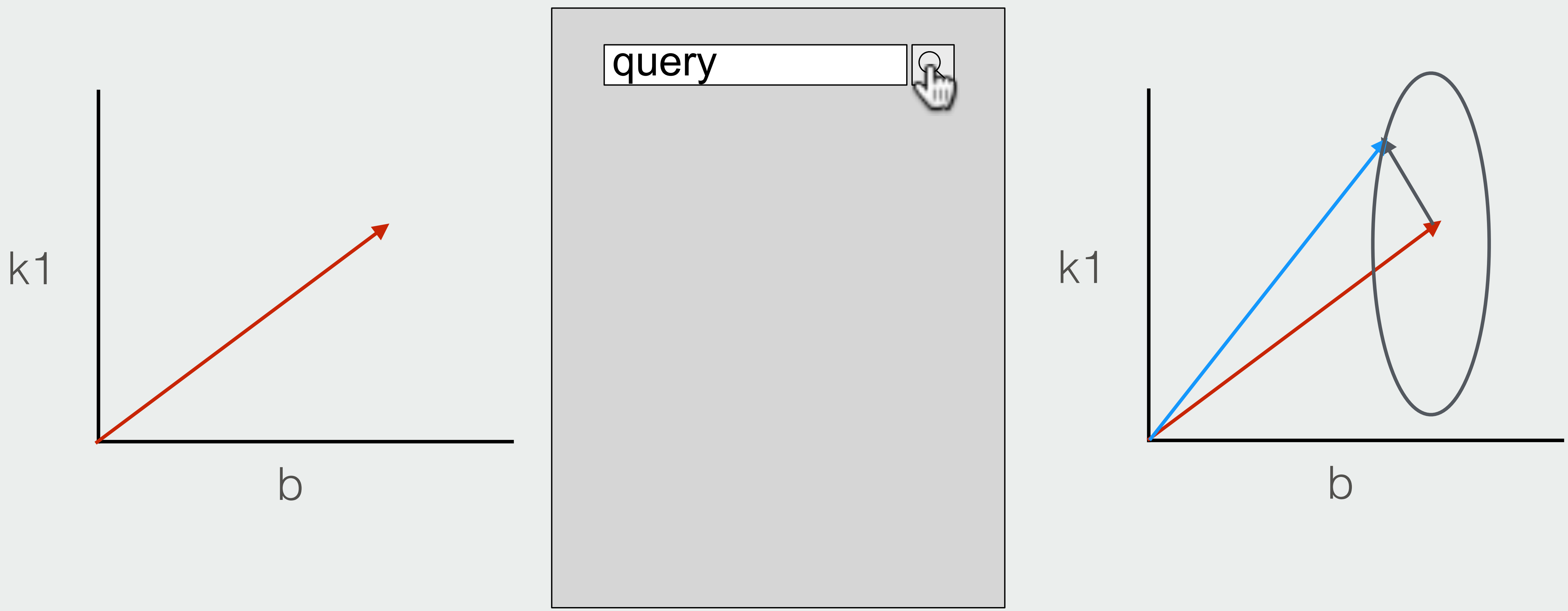




[Yue et al, 2009; Hofmann et al., 2011]

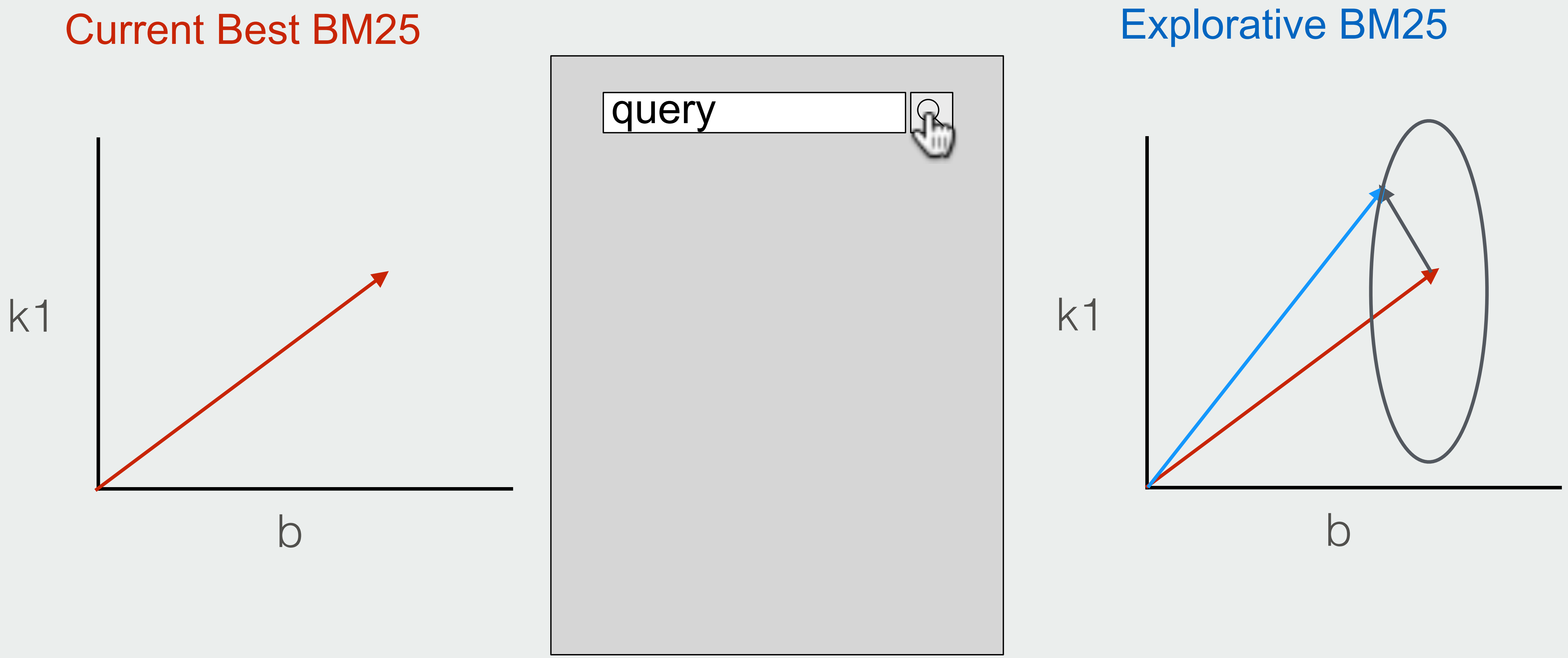
Dueling Bandit Gradient Descent

Current Best BM25



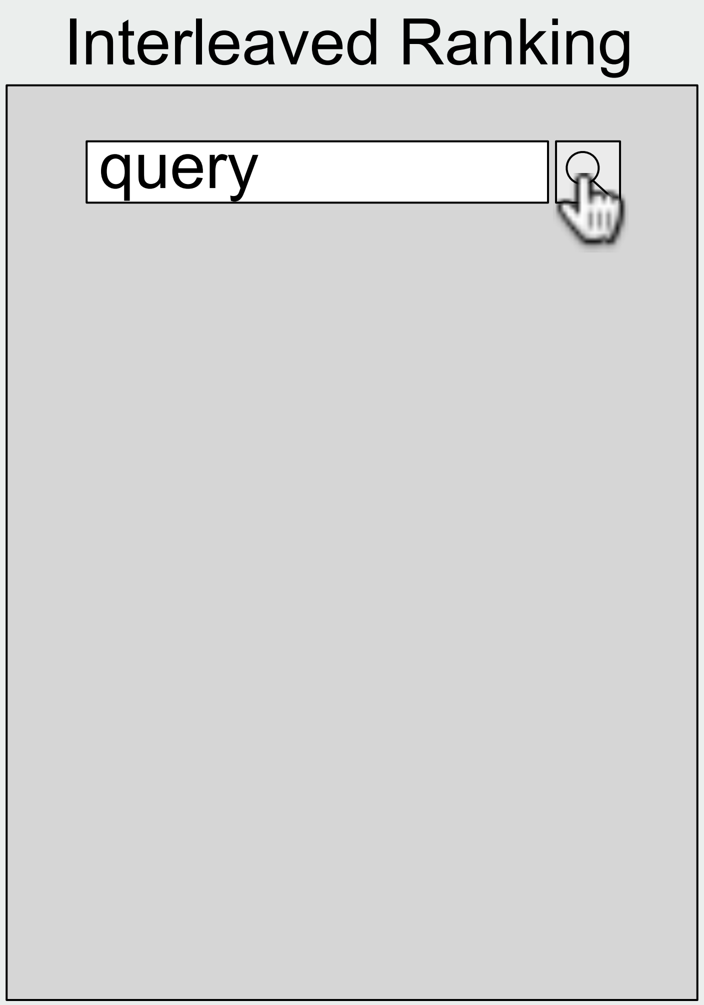
[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent



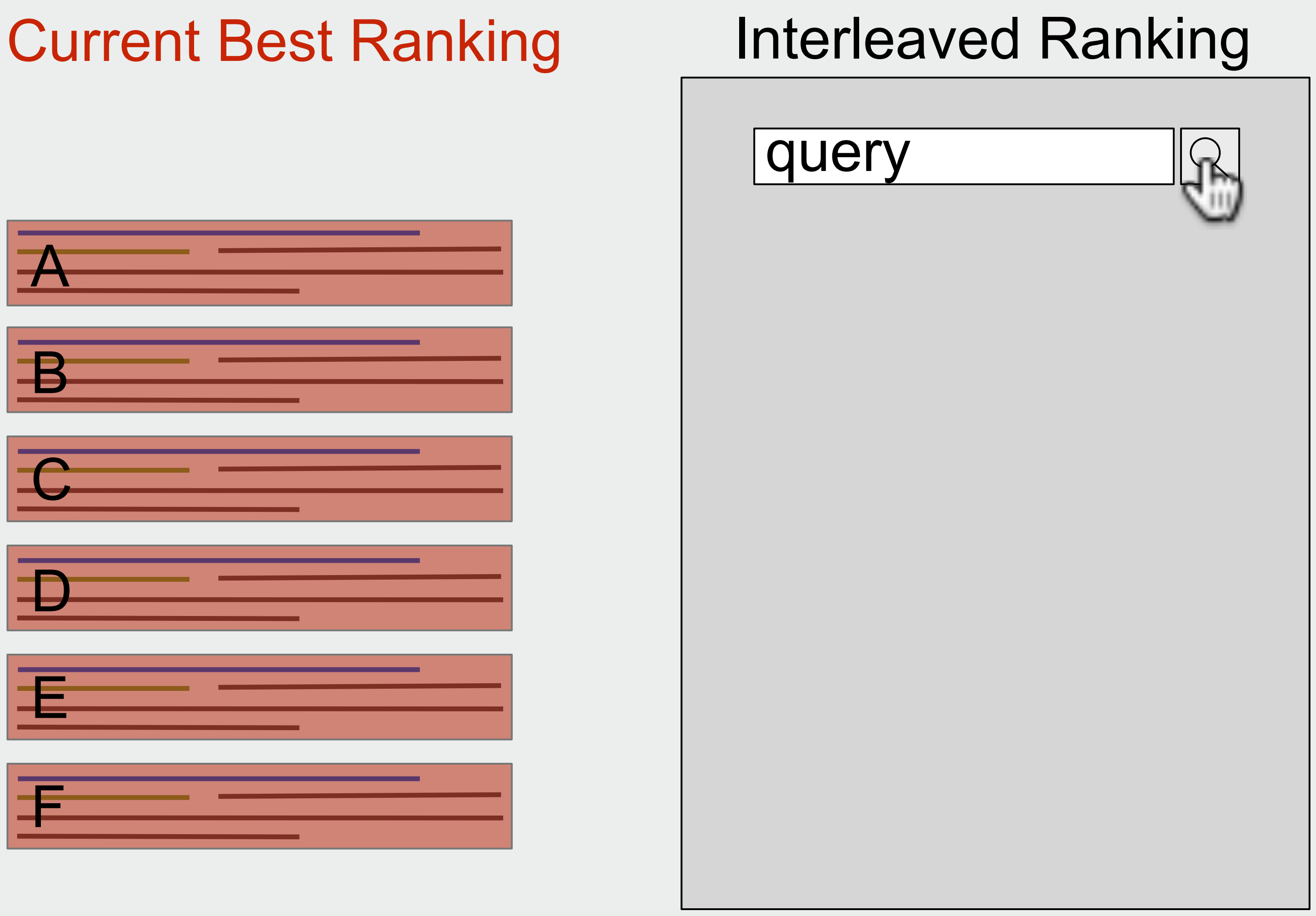
[Yue et al, 2009; Hofmann et al., 2011]

TeamDraft Interleave



[Radlinski et al.,2008]

TeamDraft Interleave

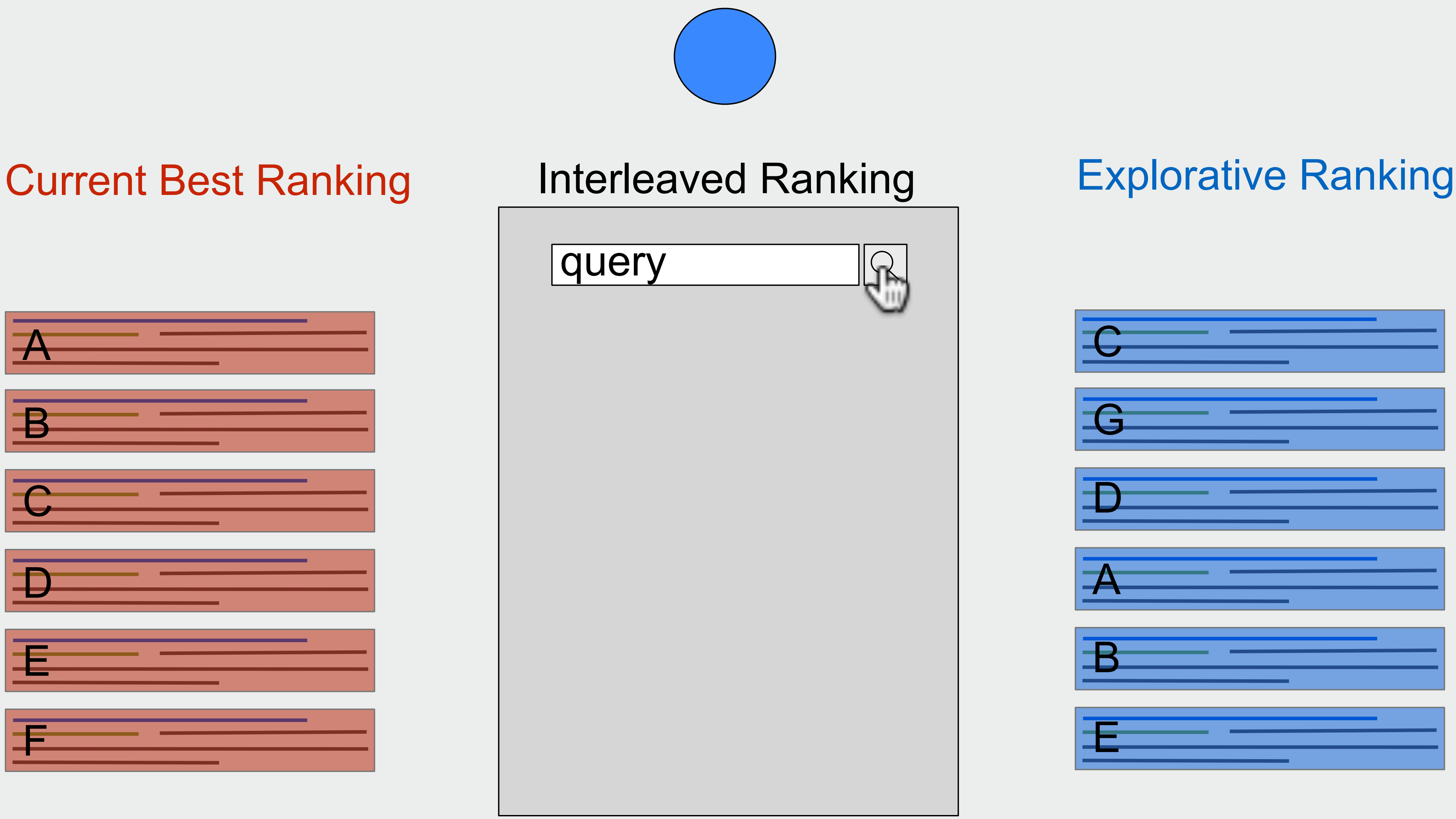


[Radlinski et al.,2008]

TeamDraft Interleave

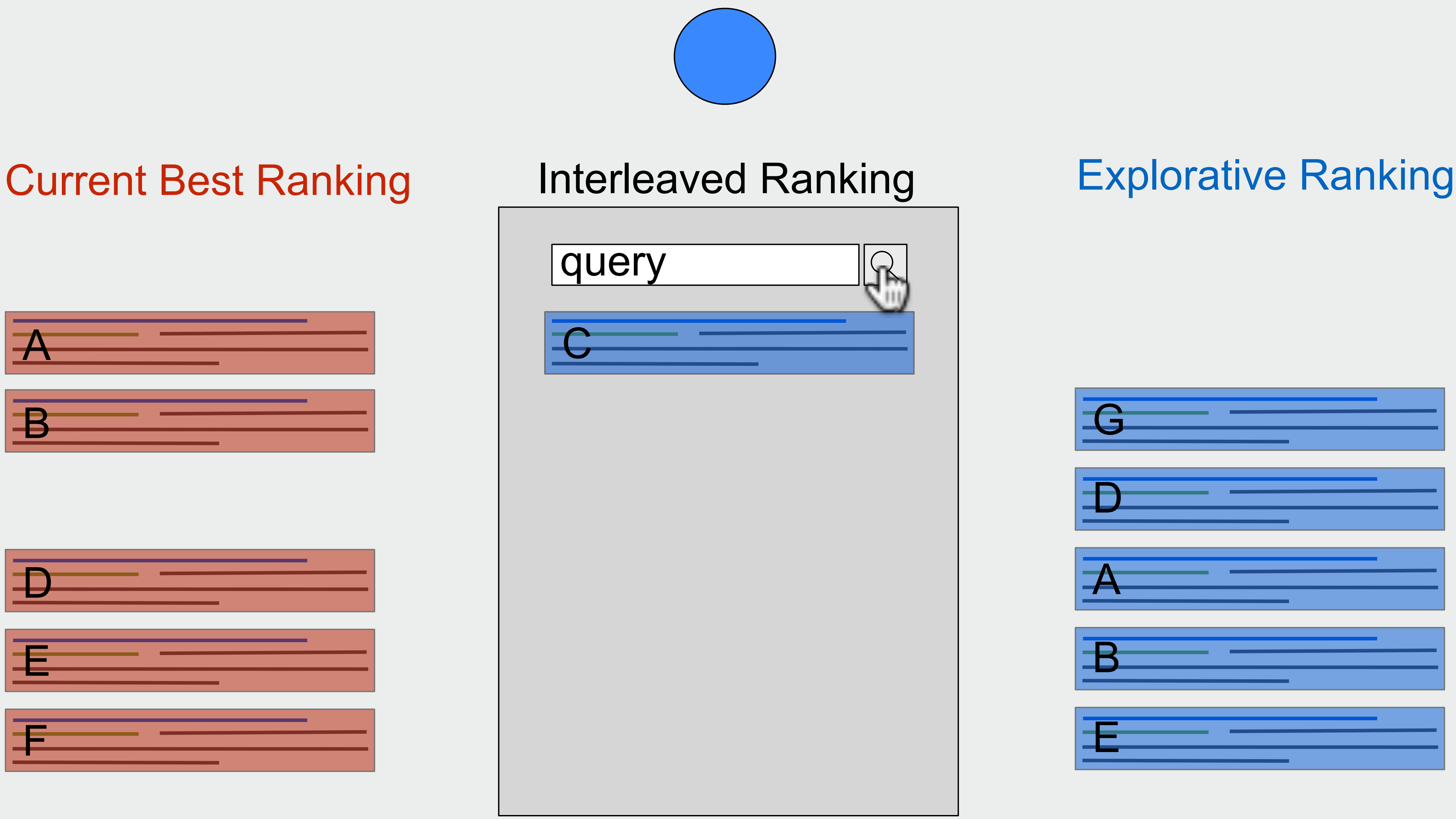
[Radlinski et al.,2008]

TeamDraft Interleave



[Radlinski et al.,2008]

TeamDraft Interleave



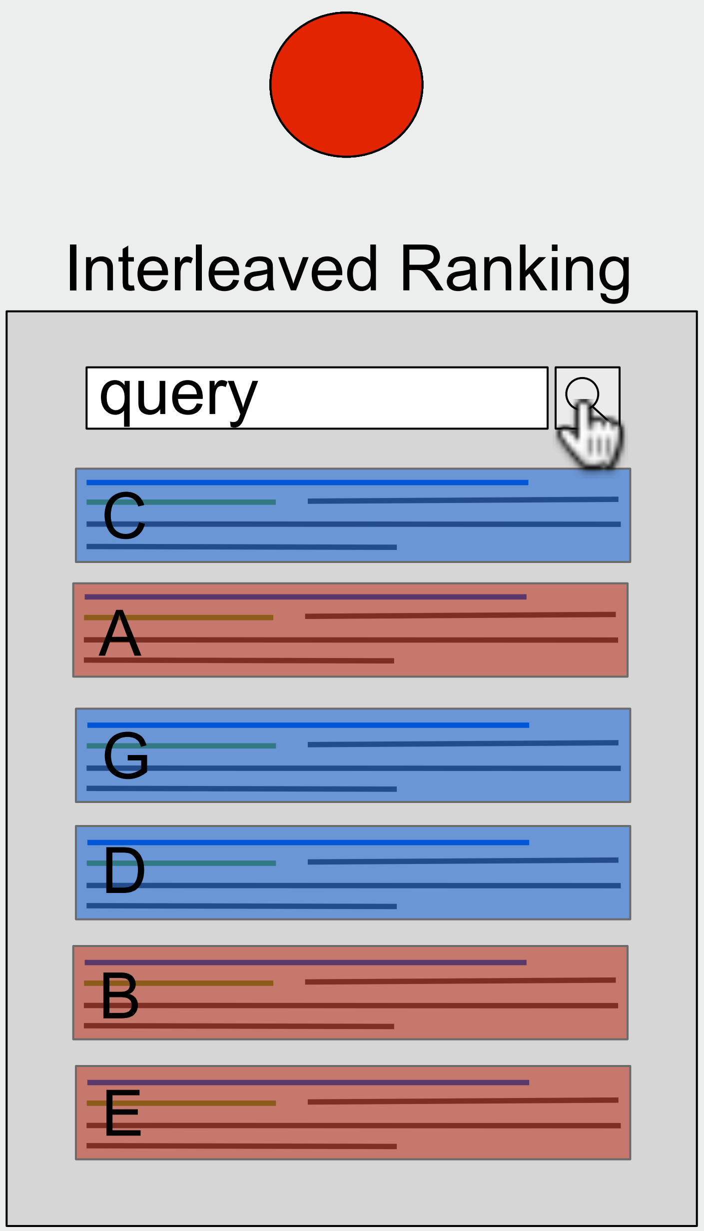
[Radlinski et al.,2008]

TeamDraft Interleave

Current Best Ranking

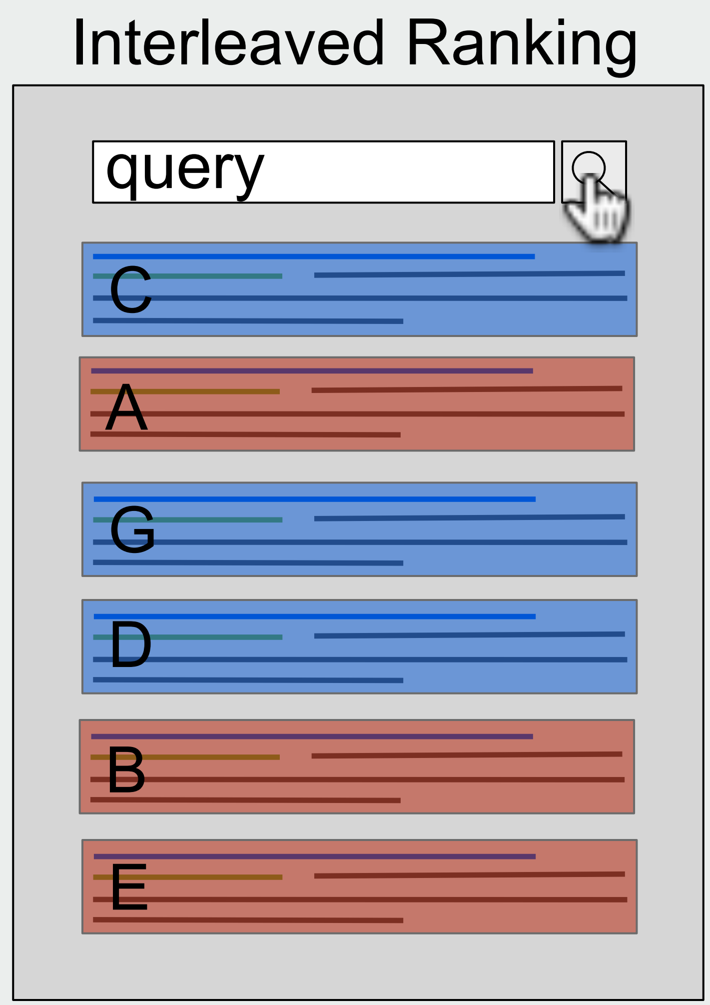
Interleaved Ranking

Explorative Ranking



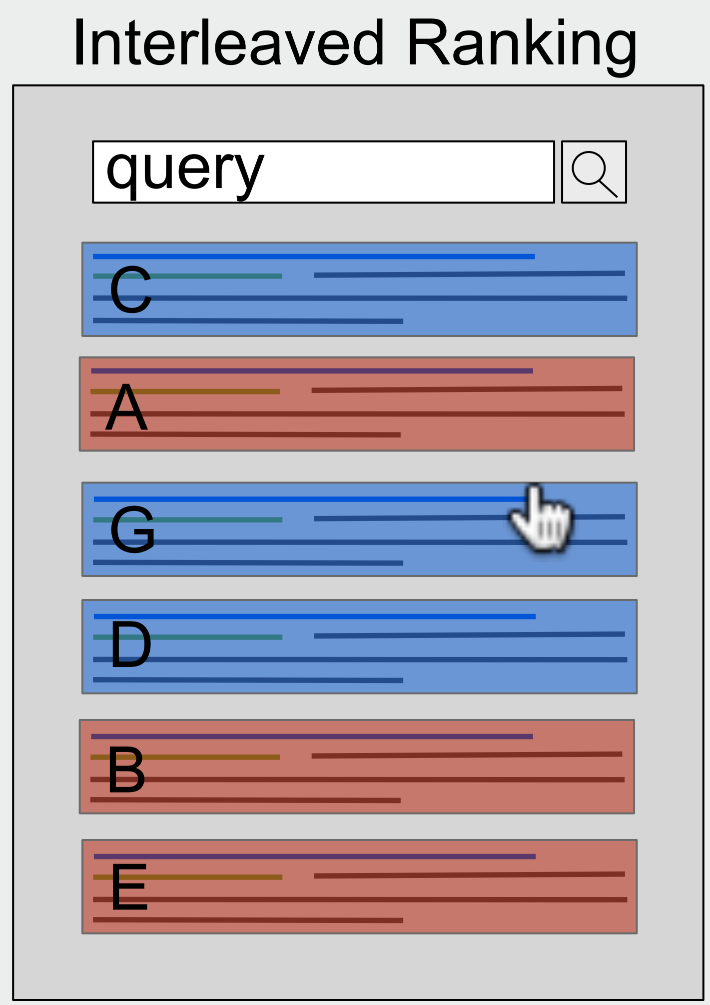
[Radlinski et al.,2008]

TeamDraft Interleave



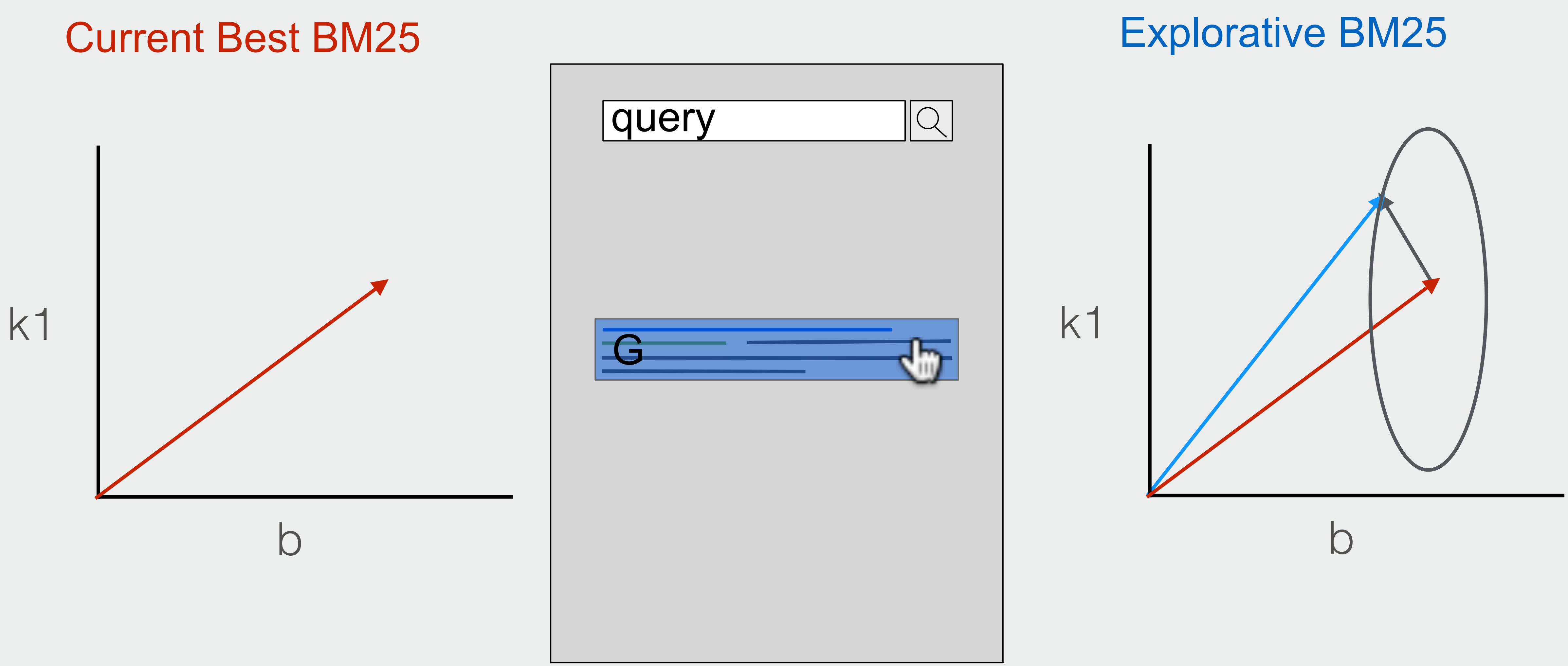
[Radlinski et al.,2008]

TeamDraft Interleave



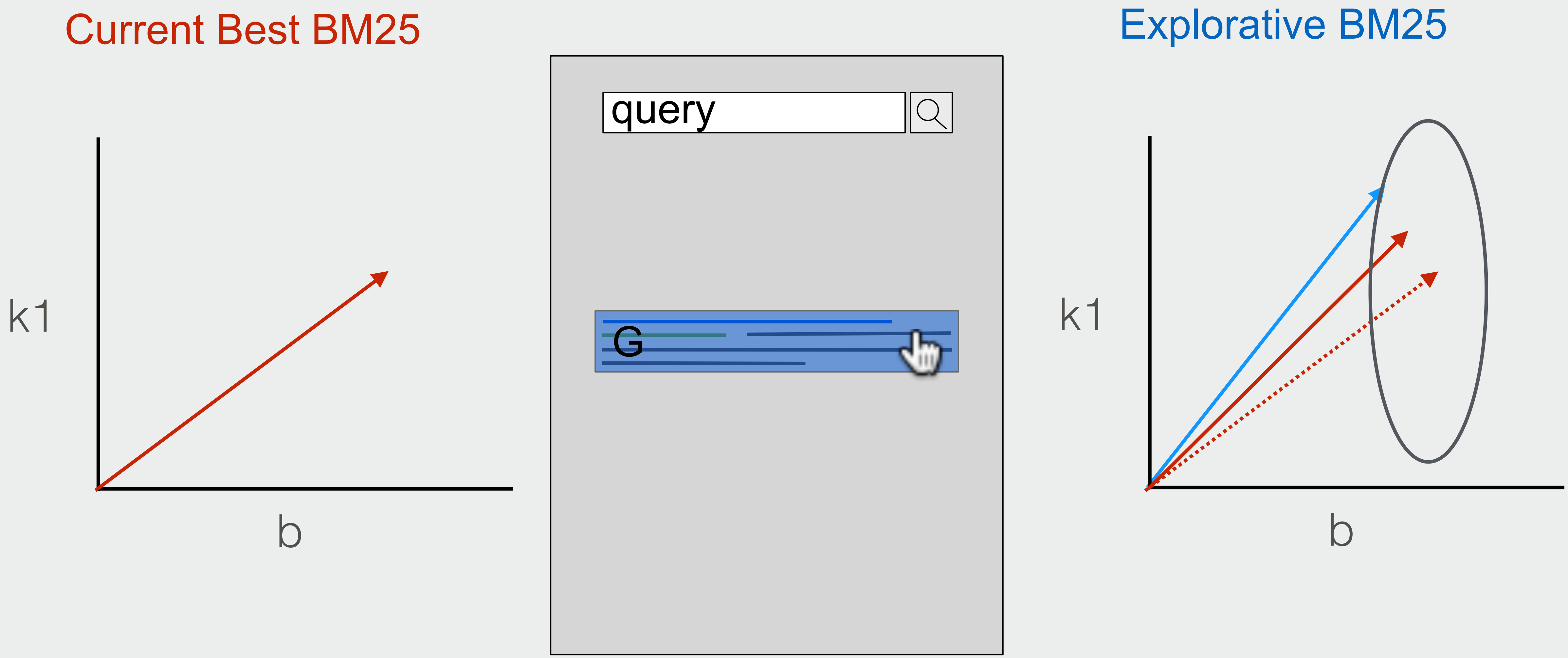
[Radlinski et al.,2008]

Dueling Bandit Gradient Descent



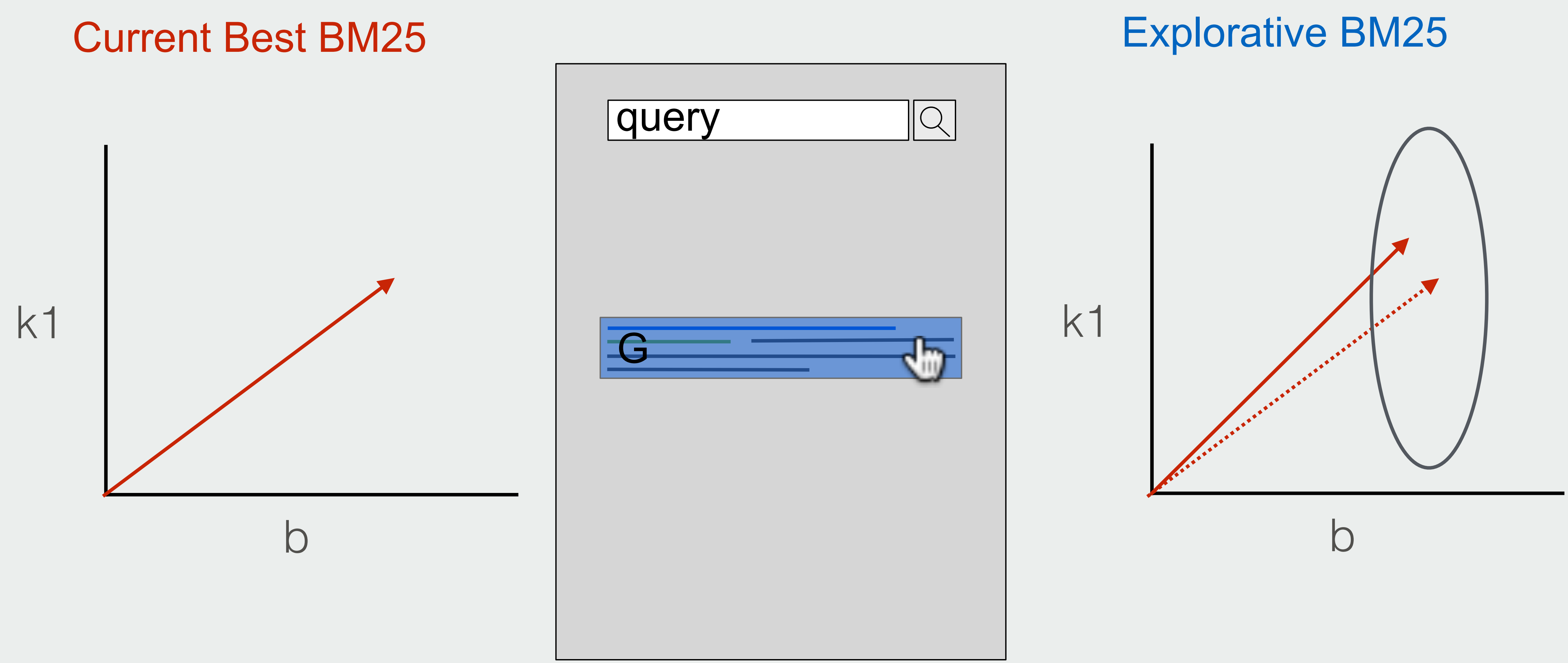
[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent



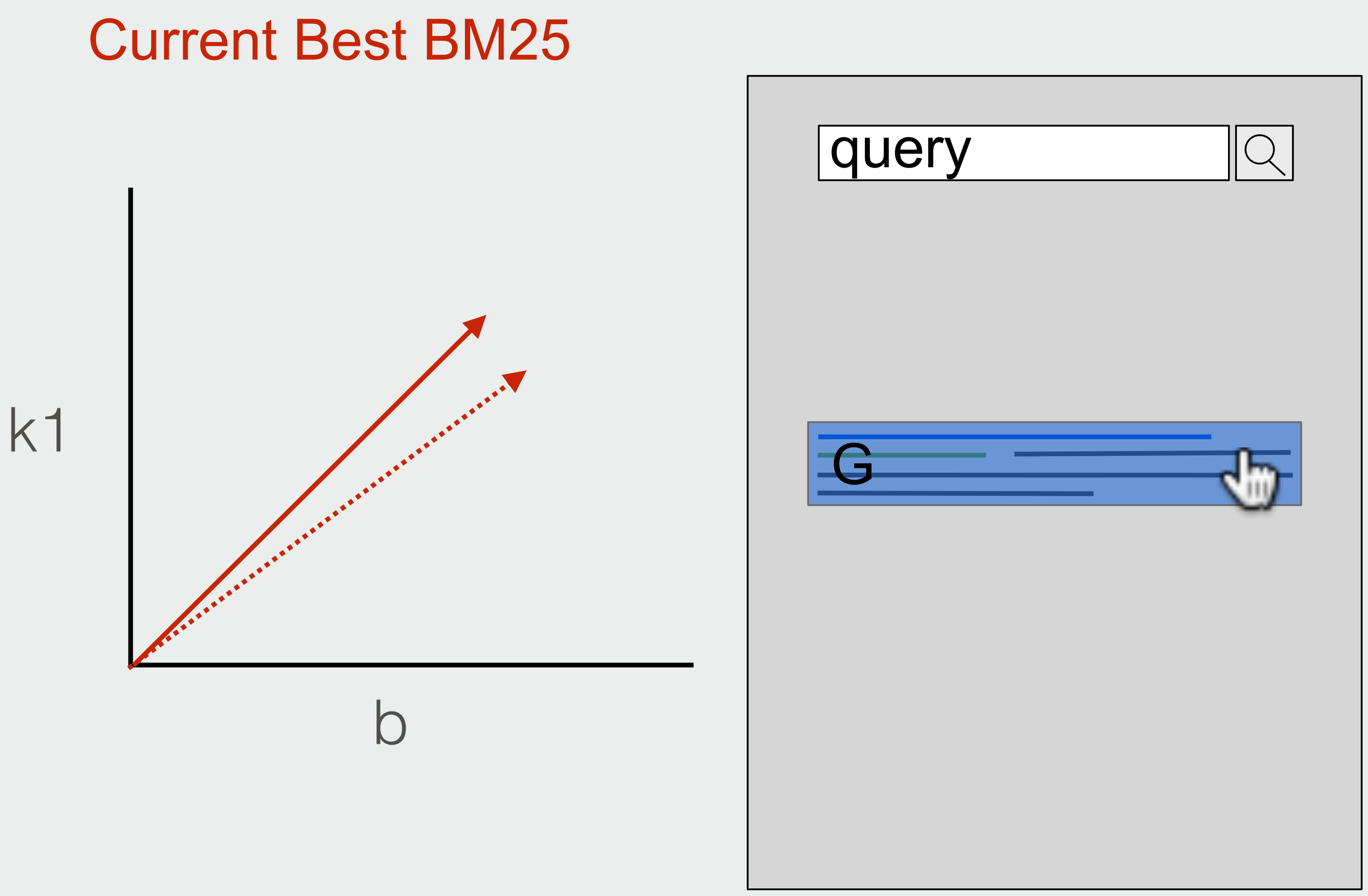
[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent



[Yue et al, 2009; Hofmann et al., 2011]

Dueling Bandit Gradient Descent



[Yue et al, 2009; Hofmann et al., 2011]

Outline

- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

Experiments

- Clicks
  - Simulated using simple click model
- Data
  - LETOR *(HP2003/4, NP2003/4, TD2003/4)*
- Software
  - Lerot bitbucket.org/ilps/lerot

UNIVERSITY OF AMSTERDAM

Outline

- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

Parameter sweep: room for improvement

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80							0.613

Parameter sweep: room for improvement

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613

Parameter sweep: room for improvement

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661▼	0.591	0.423▲	0.477	0.614

University of Amsterdam

Parameter sweep: room for improvement

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661$^{▼}$	0.591	0.423$^{▲}$	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**$^{\triangle}$	0.657$^{▼}$	0.575	0.423$^{▲}$	0.482$^{\triangle}$	0.613
	2.50	0.85	0.671	0.613	0.682	0.579$^{\triangledown}$	0.404	0.473	0.605$^{\triangledown}$
	7.30	0.80	0.690	0.647	0.661$^{▼}$	0.592	0.423$^{▲}$	0.477	0.613

Parameter sweep: room for improvement

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661$^{\blacktriangledown}$	0.591	0.423$^{\blacktriangle}$	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**$^{\vartriangle}$	0.657$^{\blacktriangledown}$	0.575	0.423$^{\blacktriangle}$	0.482$^{\vartriangle}$	0.613
	2.50	0.85	0.671	0.613	0.682	0.579$^{\triangledown}$	0.404	0.473	0.605$^{\triangledown}$
	7.30	0.80	0.690	0.647	0.661$^{\blacktriangledown}$	0.592	0.423$^{\blacktriangle}$	0.477	0.613
NP2003	2.60	0.45	0.661	0.572$^{\triangledown}$	**0.719**	0.635	0.374$^{\blacktriangledown}$	0.441$^{\blacktriangledown}$	0.607
	2.50	0.45	0.660	0.572$^{\triangledown}$	0.718	0.635	0.374$^{\blacktriangledown}$	0.441$^{\blacktriangledown}$	0.607
	2.60	0.80	0.675	0.629	0.692	0.601	0.403	0.470	0.613
NP2004	4.00	0.50	0.663	0.584	0.705	**0.647**$^{\vartriangle}$	0.386$^{\triangledown}$	0.446$^{\blacktriangledown}$	0.609
	2.50	0.50	0.663	0.573$^{\triangledown}$	0.713	0.635	0.381$^{\blacktriangledown}$	0.444$^{\blacktriangledown}$	0.607
	4.00	0.80	0.680	0.645	0.683	0.605	0.414$^{\vartriangle}$	0.474	0.616
TD2003	25.90	0.90	0.660	0.597	0.515$^{\blacktriangledown}$	0.478$^{\blacktriangledown}$	**0.456**$^{\vartriangle}$	0.489$^{\vartriangle}$	0.550$^{\blacktriangledown}$
	2.50	0.90	0.676	0.607	0.672	0.560$^{\blacktriangledown}$	0.405	0.471	0.600$^{\blacktriangledown}$
	25.90	0.80	0.645	0.576	0.535$^{\blacktriangledown}$	0.493$^{\blacktriangledown}$	0.445	0.482	0.549$^{\blacktriangledown}$
TD2004	24.00	0.90	0.664	0.604	0.520$^{\blacktriangledown}$	0.481$^{\blacktriangledown}$	0.449$^{\vartriangle}$	**0.491**$^{\vartriangle}$	0.553$^{\blacktriangledown}$
	2.50	0.90	0.676	0.607	0.672	0.560$^{\blacktriangledown}$	0.405	0.471	0.600$^{\blacktriangledown}$
	24.00	0.80	0.645	0.578	0.538$^{\blacktriangledown}$	0.496$^{\blacktriangledown}$	0.446	0.482	0.550$^{\blacktriangledown}$

Parameter sweep: room for

RQ1: How good are the manually tuned parameter values of BM25 that are currently used?

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661$^{\blacktriangledown}$	0.591	0.423$^{\blacktriangle}$	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**$^{\triangle}$	0.657$^{\blacktriangledown}$	0.575	0.423$^{\blacktriangle}$	0.482$^{\triangle}$	0.613
	2.50	0.85	0.671	0.613	0.682	0.579$^{\triangledown}$	0.404	0.473	0.605$^{\triangledown}$
	7.30	0.80	0.690	0.647	0.661$^{\blacktriangledown}$	0.592	0.423$^{\blacktriangle}$	0.477	0.613
NP2003	2.60	0.45	0.661	0.572$^{\triangledown}$	**0.719**	0.635	0.374$^{\blacktriangledown}$	0.441$^{\blacktriangledown}$	0.607
	2.50	0.45	0.660	0.572$^{\triangledown}$	0.718	0.635	0.374$^{\blacktriangledown}$	0.441$^{\blacktriangledown}$	0.607
	2.60	0.80	0.675	0.629	0.692	0.601	0.403	0.470	0.613
NP2004	4.00	0.50	0.663	0.584	0.705	**0.647**$^{\triangle}$	0.386$^{\triangledown}$	0.446$^{\blacktriangledown}$	0.609
	2.50	0.50	0.663	0.573$^{\triangledown}$	0.713	0.635	0.381$^{\blacktriangledown}$	0.444$^{\blacktriangledown}$	0.607
	4.00	0.80	0.680	0.645	0.683	0.605	0.414$^{\triangle}$	0.474	0.616
TD2003	25.90	0.90	0.660	0.597	0.515$^{\blacktriangledown}$	0.478$^{\blacktriangledown}$	**0.456**$^{\triangle}$	0.489$^{\triangle}$	0.550$^{\blacktriangledown}$
	2.50	0.90	0.676	0.607	0.672	0.560$^{\blacktriangledown}$	0.405	0.471	0.600$^{\blacktriangledown}$
	25.90	0.80	0.645	0.576	0.535$^{\blacktriangledown}$	0.493$^{\blacktriangledown}$	0.445	0.482	0.549$^{\blacktriangledown}$
TD2004	24.00	0.90	0.664	0.604	0.520$^{\blacktriangledown}$	0.481$^{\blacktriangledown}$	0.449$^{\triangle}$	**0.491**$^{\triangle}$	0.553$^{\blacktriangledown}$
	2.50	0.90	0.676	0.607	0.672	0.560$^{\blacktriangledown}$	0.405	0.471	0.600$^{\blacktriangledown}$
	24.00	0.80	0.645	0.578	0.538$^{\blacktriangledown}$	0.496$^{\blacktriangledown}$	0.446	0.482	0.550$^{\blacktriangledown}$

Parameter sweep: room for

RQ1: How good are the manually tuned parameter values of BM25 that are currently used?

quite good

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661▼	0.591	0.423▲	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**△	0.657▼	0.575	0.423▲	0.482△	0.613
	2.50	0.85	0.671	0.613	0.682	0.579▽	0.404	0.473	0.605▽
	7.30	0.80	0.690	0.647	0.661▼	0.592	0.423▲	0.477	0.613
NP2003	2.60	0.45	0.661	0.572▽	**0.719**	0.635	0.374▼	0.441▼	0.607
	2.50	0.45	0.660	0.572▽	0.718	0.635	0.374▼	0.441▼	0.607
	2.60	0.80	0.675	0.629	0.692	0.601	0.403	0.470	0.613
NP2004	4.00	0.50	0.663	0.584	0.705	**0.647**△	0.386▽	0.446▼	0.609
	2.50	0.50	0.663	0.573▽	0.713	0.635	0.381▼	0.444▼	0.607
	4.00	0.80	0.680	0.645	0.683	0.605	0.414△	0.474	0.616
TD2003	25.90	0.90	0.660	0.597	0.515▼	0.478▼	**0.456**△	0.489△	0.550▼
	2.50	0.90	0.676	0.607	0.672	0.560▼	0.405	0.471	0.600▼
	25.90	0.80	0.645	0.576	0.535▼	0.493▼	0.445	0.482	0.549▼
TD2004	24.00	0.90	0.664	0.604	0.520▼	0.481▼	0.449△	**0.491**△	0.553▼
	2.50	0.90	0.676	0.607	0.672	0.560▼	0.405	0.471	0.600▼
	24.00	0.80	0.645	0.578	0.538▼	0.496▼	0.446	0.482	0.550▼

LiMo SINe

Parameter sweep: room for

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661▼	0.591	0.423▲	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**△	0.657▼	0.575	0.423▲	0.482△	0.613
	2.50	0.85	0.671	0.613	0.682	0.579▽	0.404	0.473	0.605▽
	7.30	0.80	0.690	0.647	0.661▼	0.592	0.423▲	0.477	0.613
NP2003	2.60	0.45	0.661	0.572▽	**0.719**	0.635	0.374▼	0.441▼	0.607
	2.50	0.45	0.660	0.572▽	0.718	0.635	0.374▼	0.441▼	0.607
	2.60	0.80	0.675	0.629	0.692	0.601	0.403	0.470	0.613
NP2004	4.00	0.50	0.663	0.584	0.705	**0.647**△	0.386▽	0.446▼	0.609
	2.50	0.50	0.663	0.573▽	0.713	0.635	0.381▼	[illegible]	[illegible]
	4.00	0.80	0.680	0.645	0.683	0.605	0.414△	[illegible]	[illegible]
TD2003	25.90	0.90	0.660	0.597	0.515▼	0.478▼	**0.456**△	[illegible]	[illegible]
	2.50	0.90	0.676	0.607	0.672	0.560▼	0.405	[illegible]	[illegible]
	25.90	0.80	0.645	0.576	0.535▼	0.493▼	0.445	[illegible]	[illegible]
TD2004	24.00	0.90	0.664	0.604	0.520▼	0.481▼	0.449	[illegible]	[illegible]
	2.50	0.90	0.676	0.607	0.672	0.560▼	0.405	[illegible]	[illegible]
	24.00	0.80	0.645	0.578	0.538▼	0.496▼	0.446	0.482	0.550

RQ1: How good are the manually tuned parameter values of BM25 that are currently used?

quite good

RQ2: Are they optimal for all data sets on average? Are they optimal for individual data sets?

Parameter sweep: room for

	k_1	b	HP2003	HP2004	NP2003	NP2004	TD2003	TD2004	Overall
.gov	2.50	0.80	0.674	0.629	0.693	0.599	0.404	0.469	0.613
HP2003	7.40	0.80	**0.692**	0.650	0.661$^\blacktriangledown$	0.591	0.423$^\blacktriangle$	0.477	0.614
HP2004	7.30	0.85	0.688	**0.672**$^\vartriangle$	0.657$^\blacktriangledown$	0.575	0.423$^\blacktriangle$	0.482$^\vartriangle$	0.613
	2.50	0.85	0.671	0.613	0.682	0.579$^\triangledown$	0.404	0.473	0.605$^\triangledown$
	7.30	0.80	0.690	0.647	0.661$^\blacktriangledown$	0.592	0.423$^\blacktriangle$	0.477	0.613
NP2003	2.60	0.45	0.661	0.572$^\triangledown$	**0.719**	0.635	0.374$^\blacktriangledown$	0.441$^\blacktriangledown$	0.607
	2.50	0.45	0.660	0.572$^\triangledown$	0.718	0.635	0.374$^\blacktriangledown$	0.441$^\blacktriangledown$	0.607
	2.60	0.80	0.675	0.629	0.692	0.601	0.403	0.470	0.613
NP2004	4.00	0.50	0.663	0.584	0.705	**0.647**$^\vartriangle$	0.386$^\triangledown$	[illegible]	0.609
	2.50	0.50	0.663	0.573$^\triangledown$	0.713	0.635	0.381$^\blacktriangledown$	[illegible]	[illegible]
	4.00	0.80	0.680	0.645	0.683	0.605	0.414$^\vartriangle$	[illegible]	[illegible]
TD2003	25.90	0.90	0.660	0.597	0.515$^\blacktriangledown$	0.478$^\blacktriangledown$	**0.456**	[illegible]	[illegible]
	2.50	0.90	0.676	0.607	0.672	0.560$^\blacktriangledown$	0.405	[illegible]	[illegible]
	25.90	0.80	0.645	0.576	0.535$^\blacktriangledown$	0.493$^\blacktriangledown$	0.445	[illegible]	[illegible]
TD2004	24.00	0.90	0.664	0.604	0.520$^\blacktriangledown$	0.481$^\blacktriangledown$	0.449	[illegible]	[illegible]
	2.50	0.90	0.676	0.607	0.672	0.560$^\blacktriangledown$	0.405	[illegible]	[illegible]
	24.00	0.80	0.645	0.578	0.538$^\blacktriangledown$	0.496$^\blacktriangledown$	0.446	0.482	0.550

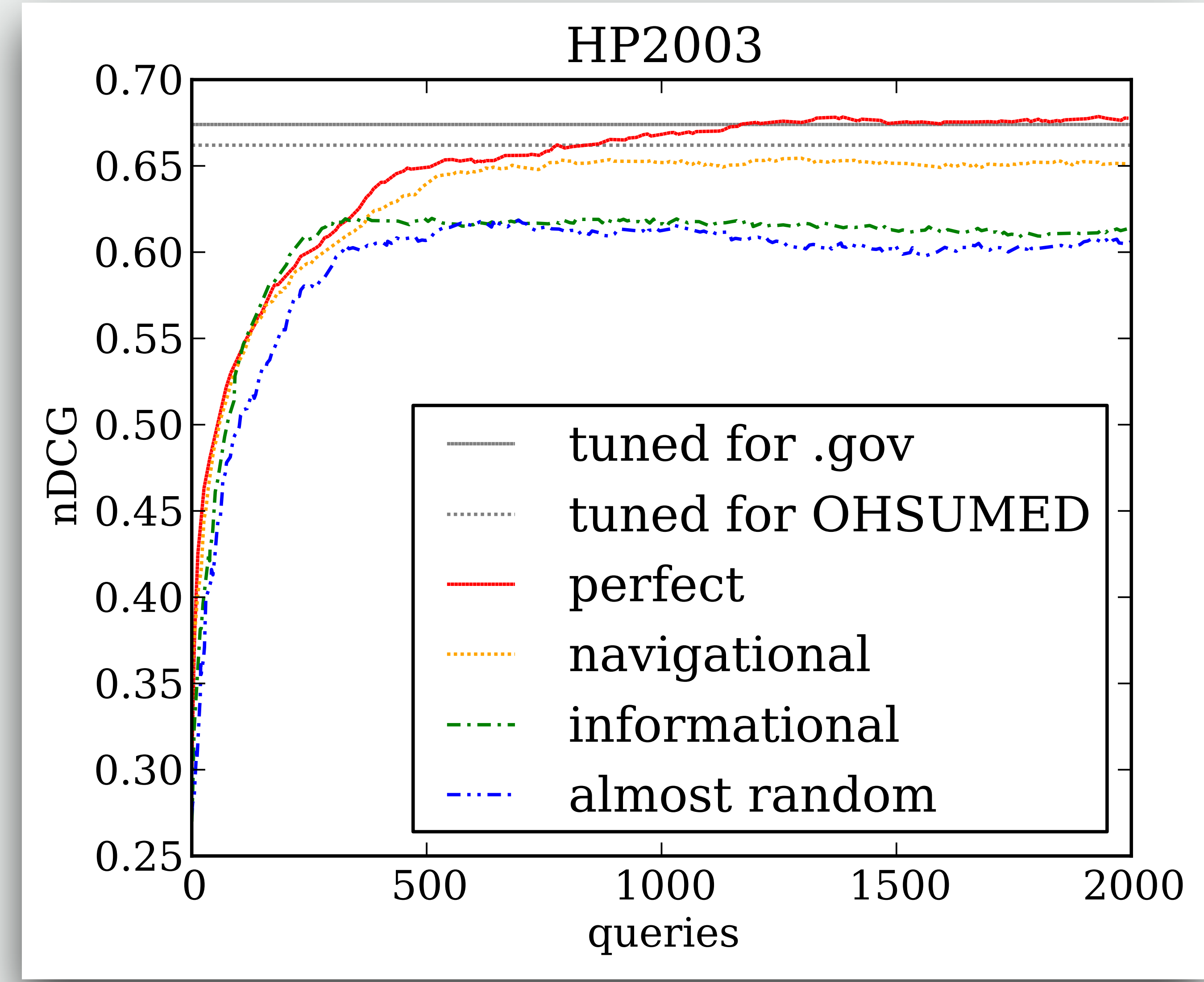
RQ1: How good are the manually tuned parameter values of BM25 that are currently used?

quite good

RQ2: Are they optimal for all data sets on average? Are they optimal for individual data sets?

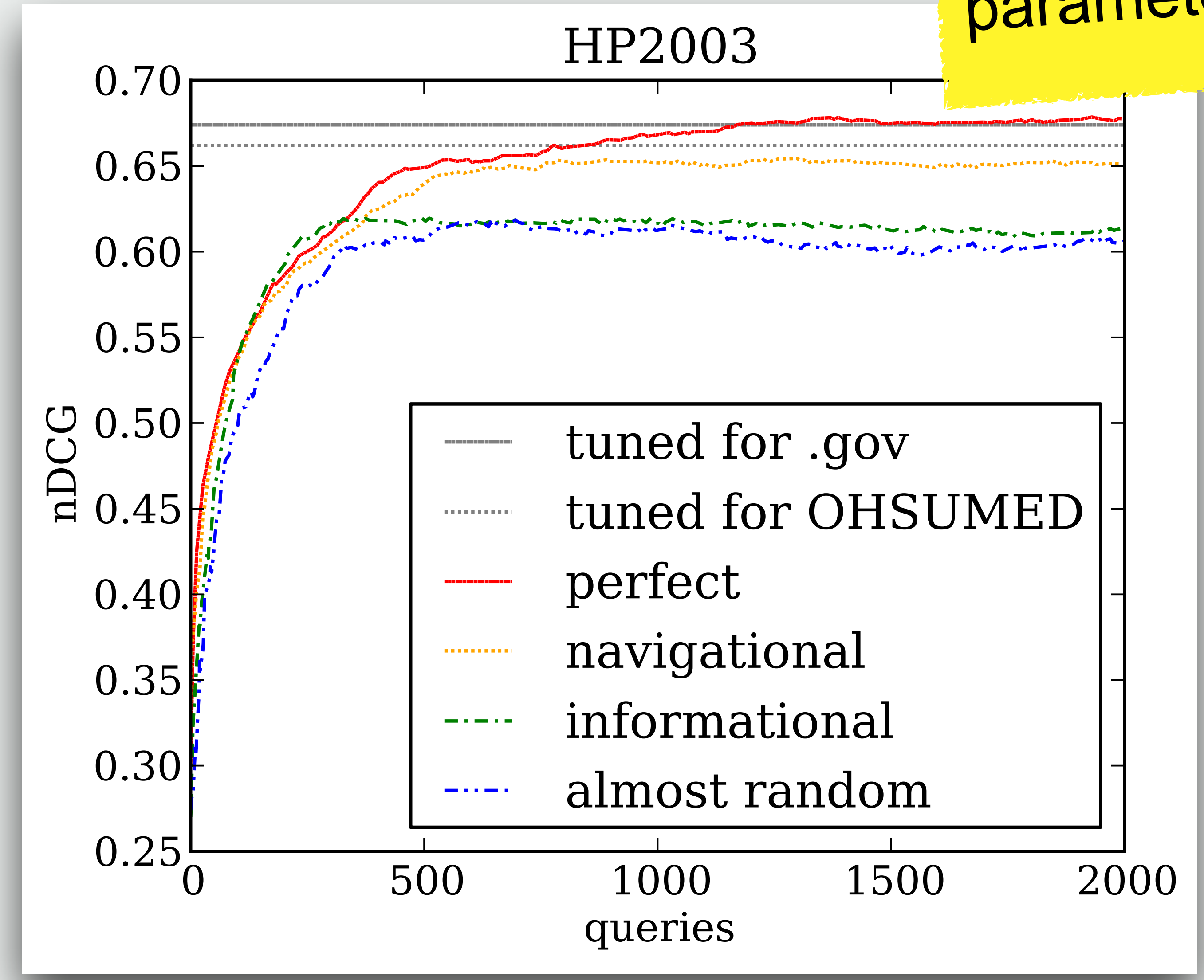
no

Learning curves



Learning curves

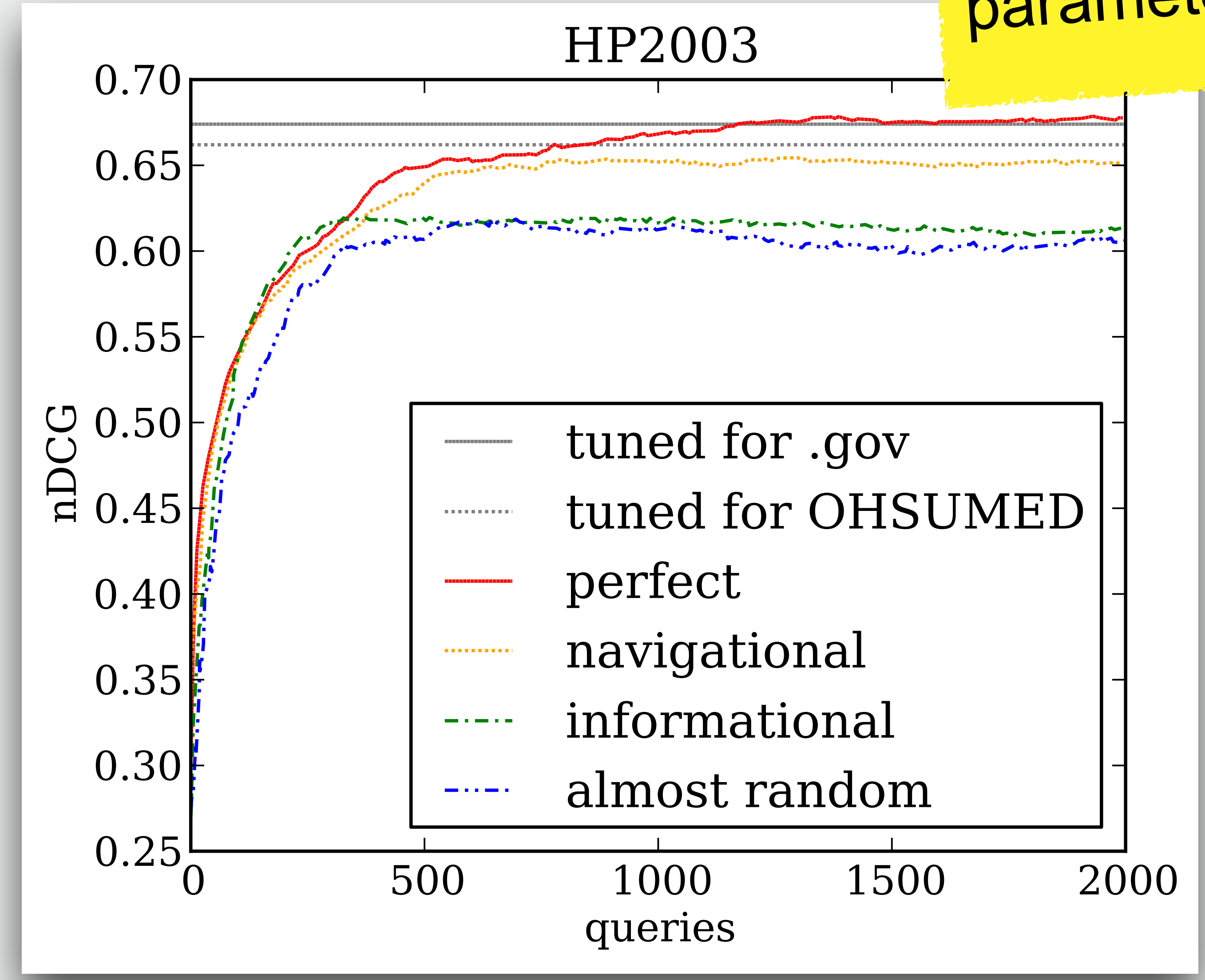
RQ3: Is it possible to learn good values of the BM25 parameters from clicks?



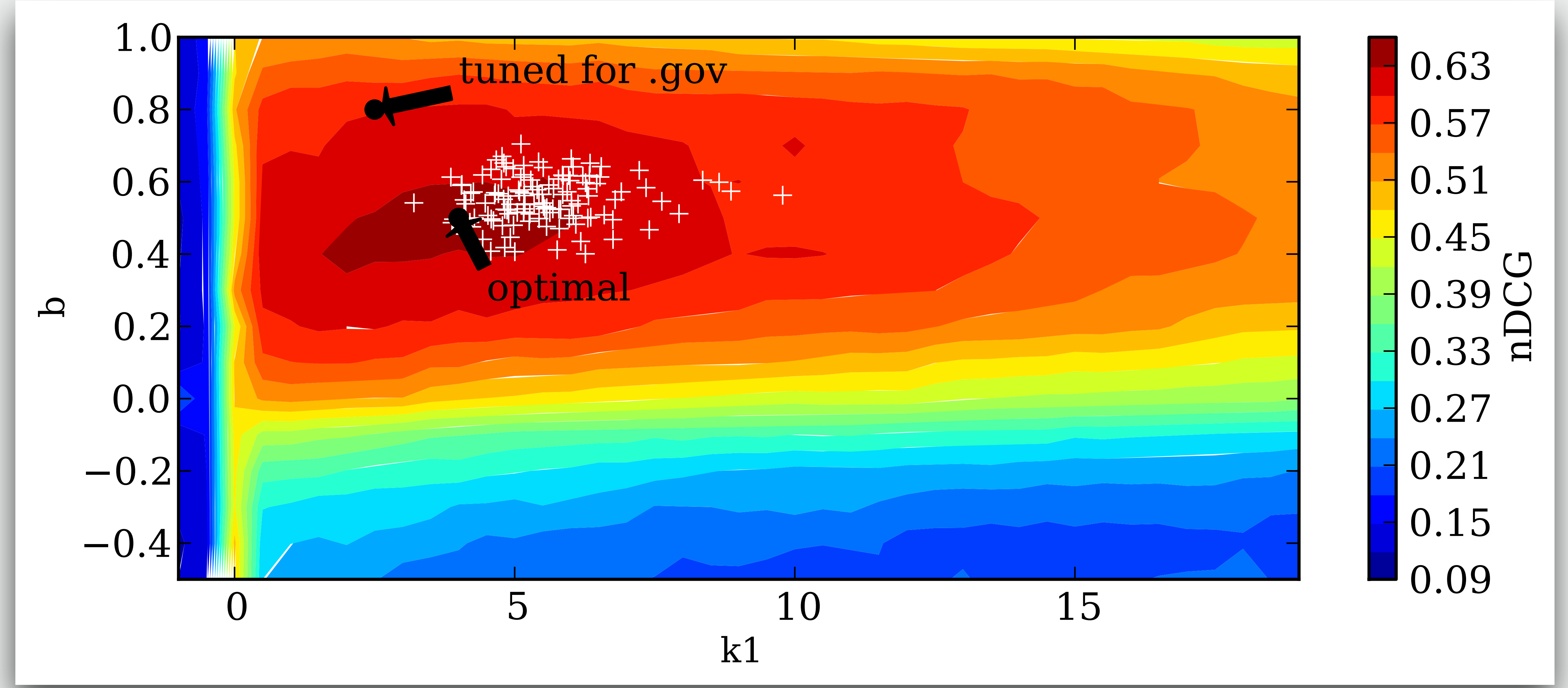
Learning curves

RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

yes

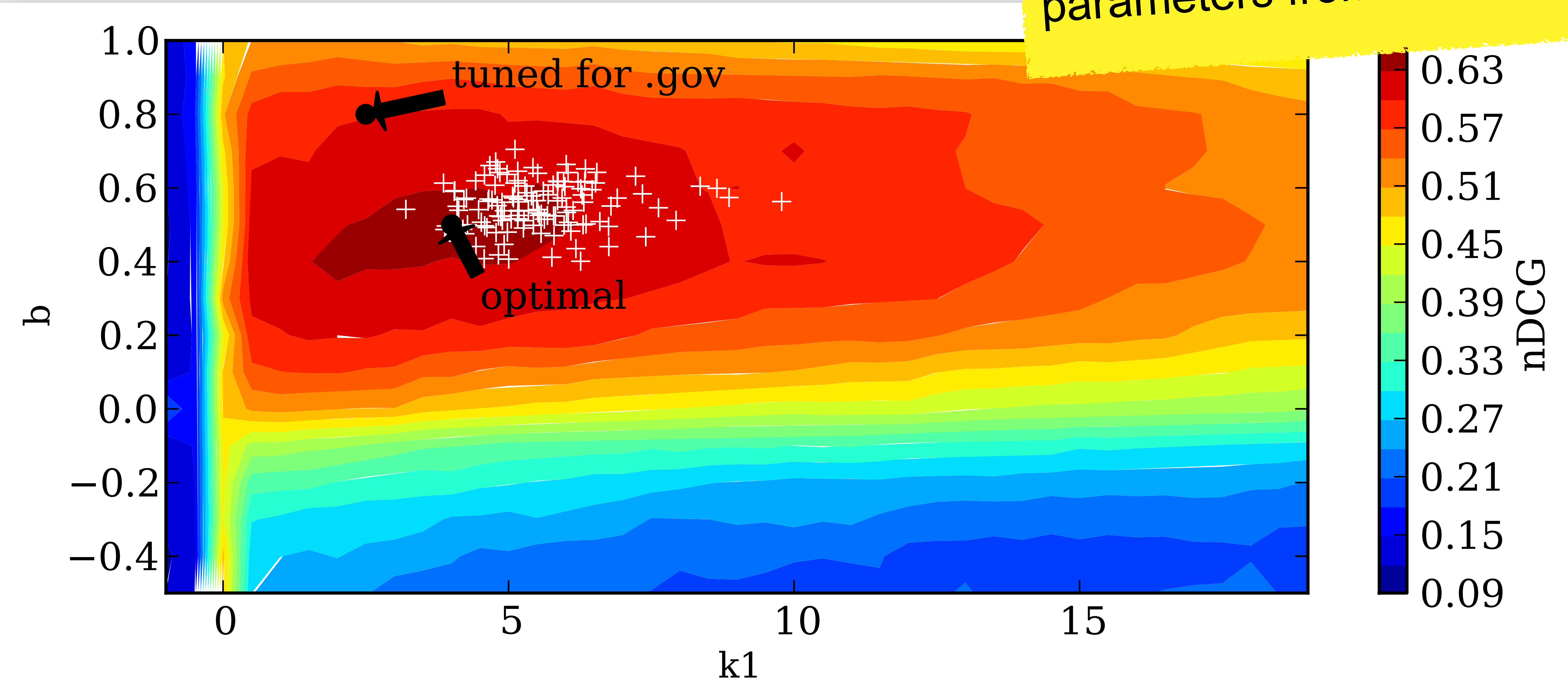


Optimization Landscape



Optimization Landscape

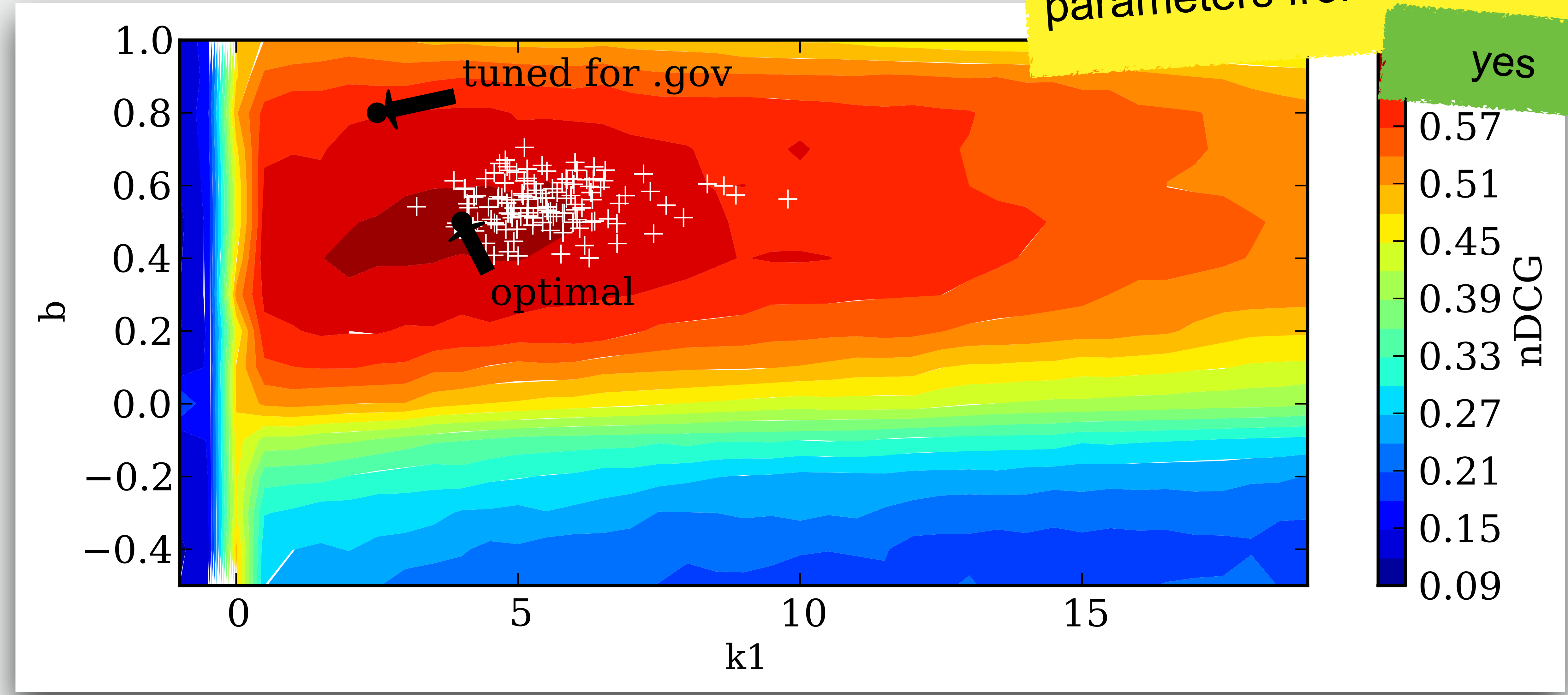
RQ3: Is it possible to learn good values of the BM25 parameters from clicks?



Optimization Landscape

RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

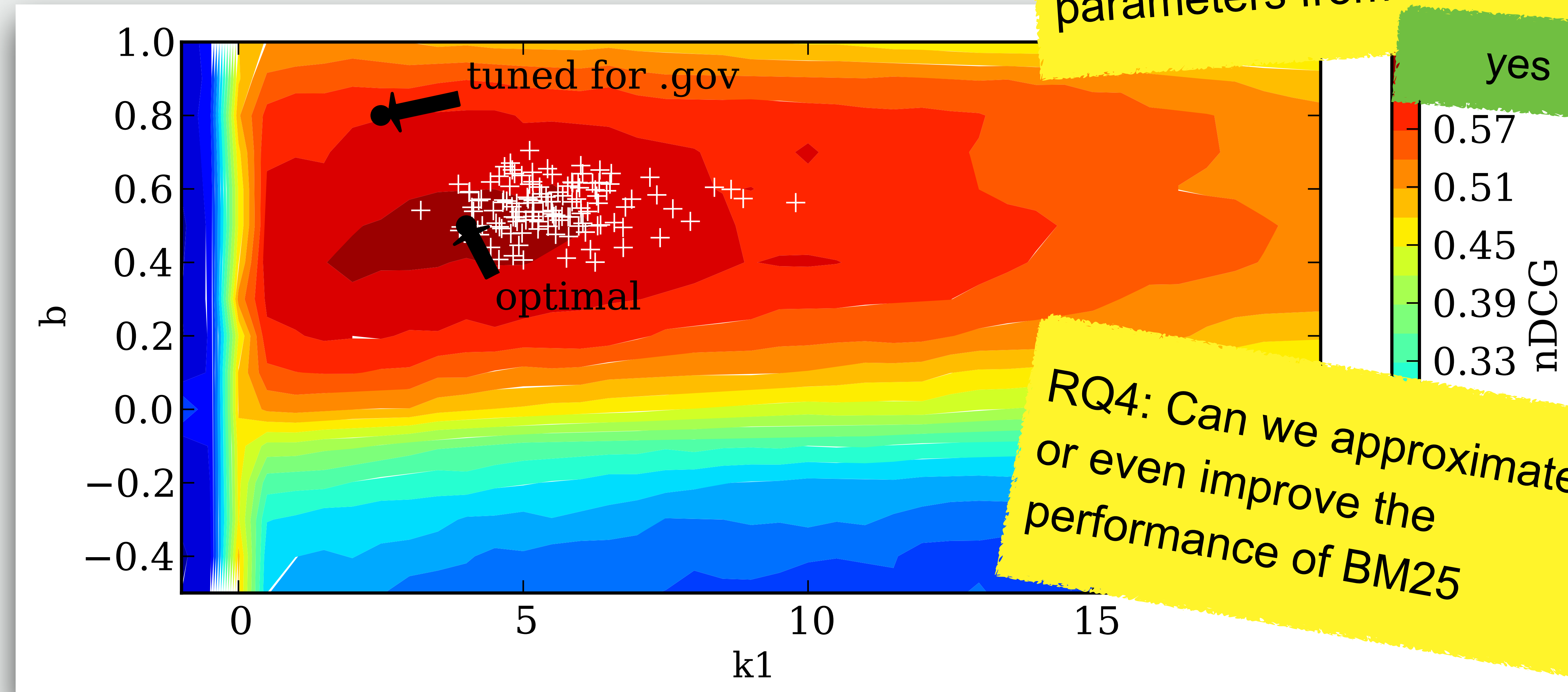
yes



Optimization Landscape

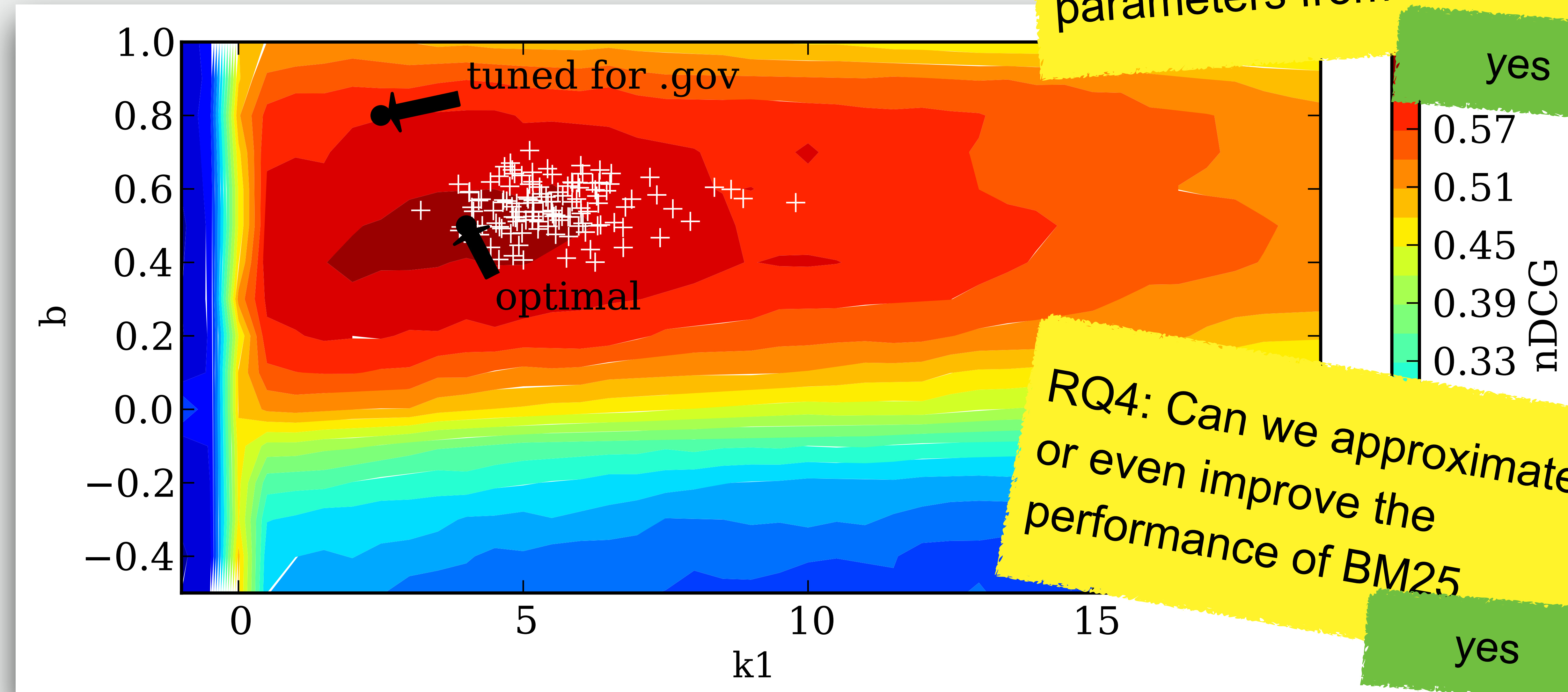
RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

yes



RQ4: Can we approximate or even improve the performance of BM25

Optimization Landscape



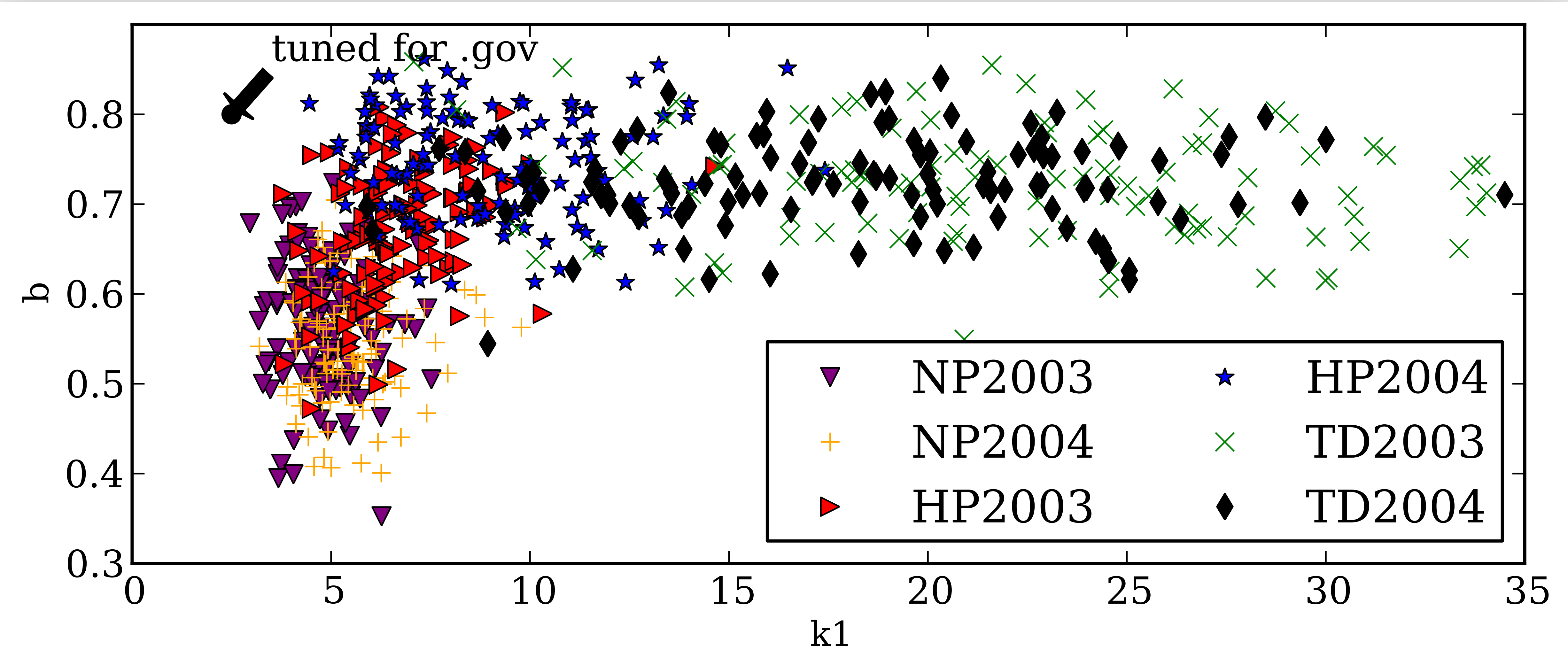
RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

yes

RQ4: Can we approximate or even improve the performance of BM25

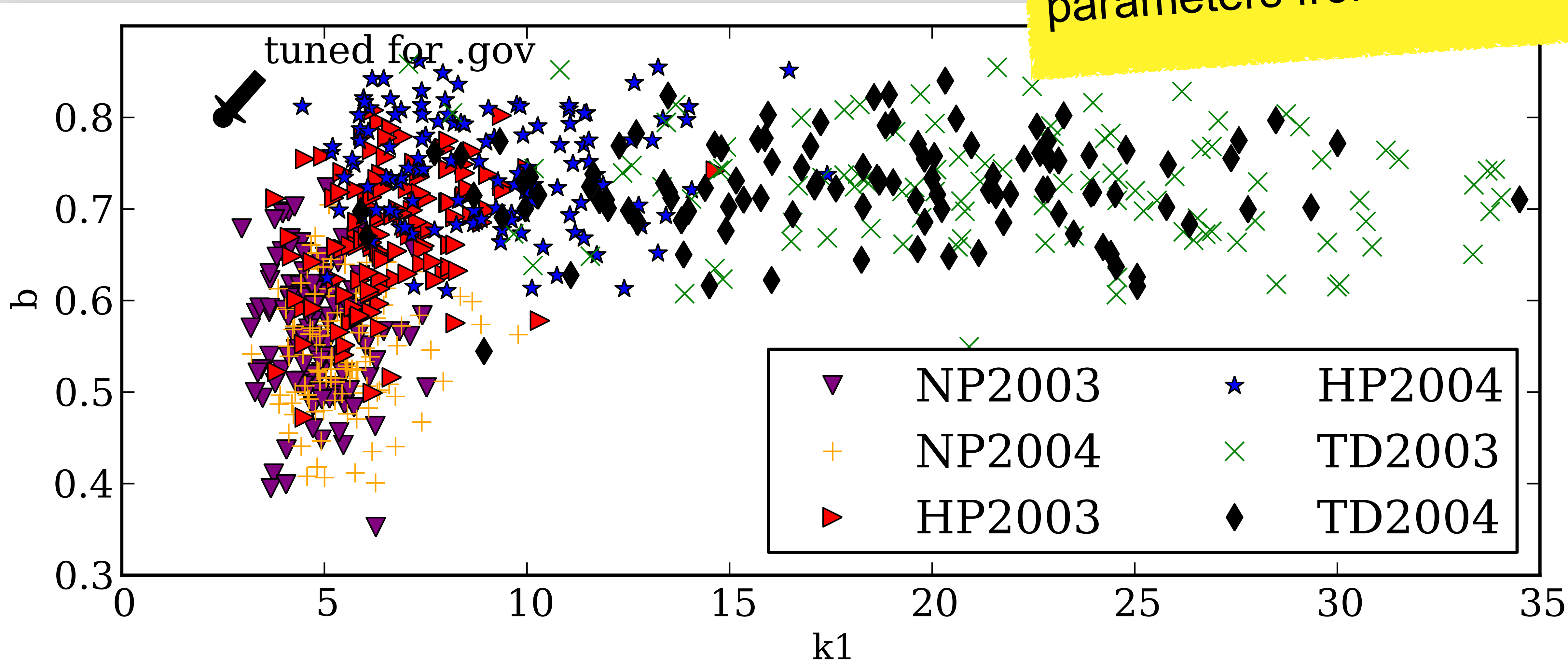
yes

Convergence per Dataset



Convergence per Dataset

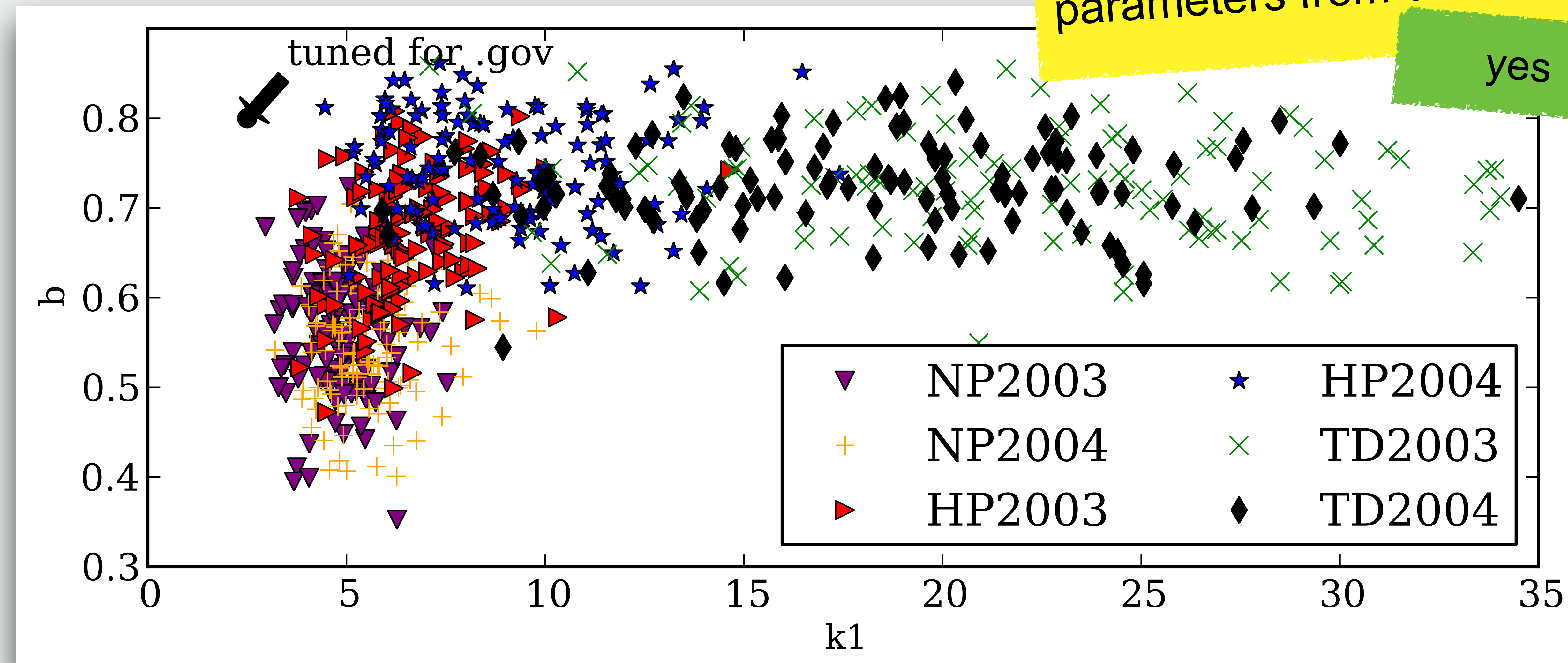
RQ3: Is it possible to learn good values of the BM25 parameters from clicks?



Convergence per Dataset

RQ3: Is it possible to learn good values of the BM25 parameters from clicks?

yes



Outline

- Introduction
  - Learning to Rank
  - Two Issues
  - Their solutions
- Research Questions
- Method
- Experiments
- Results
- Conclusions

Conclusions

Conclusions

- Parameters of base rankers matter

Conclusions

- Parameters of base rankers matter
  - values from literature are good in general but not for particular settings

Conclusions

- Parameters of base rankers matter
  - values from literature are good in general but not for particular settings
- We can learn good parameters

Conclusions

- Parameters of base rankers matter
  - values from literature are good in general but not for particular settings
- We can learn good parameters
  - from clicks

Conclusions

- Parameters of base rankers matter
  - values from literature are good in general but not for particular settings
- We can learn good parameters
  - from clicks
  - using an online algorithm

Conclusions

- Parameters of base rankers matter
  - values from literature are good in general but not for particular settings
- We can learn good parameters
  - from clicks
  - using an online algorithm
- Expensive assessments are not needed to find good parameters

thank you