

Task-aware query recommendation

Henry Feild

James Allan

Center for Intelligent Information Retrieval
University of Massachusetts Amherst

*July 29, 2013
Dublin, Ireland*



when was the dsec created



[Delaware Solar Energy Coalition](#)

The Delaware Solar Energy Coalition (**DSEC**) was founded...

<http://delsec.org>

[Deaf Smith Electric Cooperative – Home](#)

The Deaf Smith Electric Cooperative (**DSEC**).

<http://dsec.org>

Related searches:

[dsec info](#)

[dsec](#)

[delaware solar energy coalition created](#)

[deaf smith electric cooperative created](#)

[dayton service engineers collaborative created](#)

what is the dupont science essay contest



when was the dsec created



[DuPont Challenge](#)

The **DuPont** Challenge calls on students to research, think critically...

<http://thechallenge.dupont.com>

[DuPont](#)

For more than 200 years, **DuPont** has brought world-class science and engineering...

Related searches:

[dupont challenge information](#)

[dupont essay submission](#)

[dupont challenge history](#)

[deaf smith electric cooperative created](#)

[dayton service engineers collaborative created](#)

Potential benefits of context

Find key terms/phrases

elliptical trainer

elliptical trainer benefits

crisis intervention professional services

crisis intervention services

Disambiguation

object management group

andrew watson

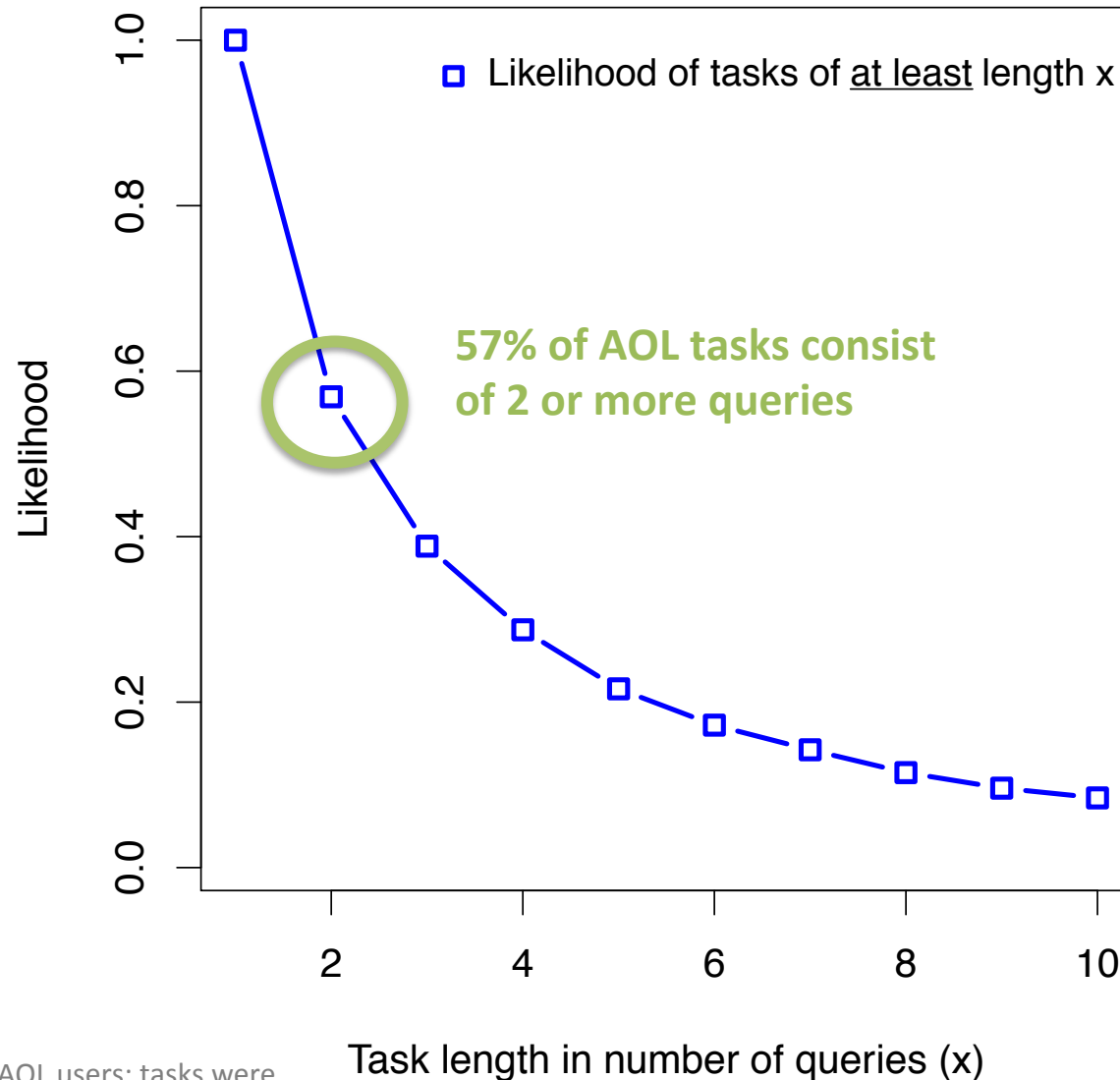
watson omg

Term expansion

michworks

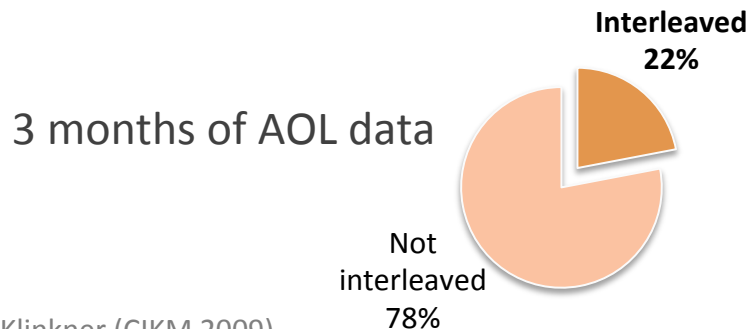
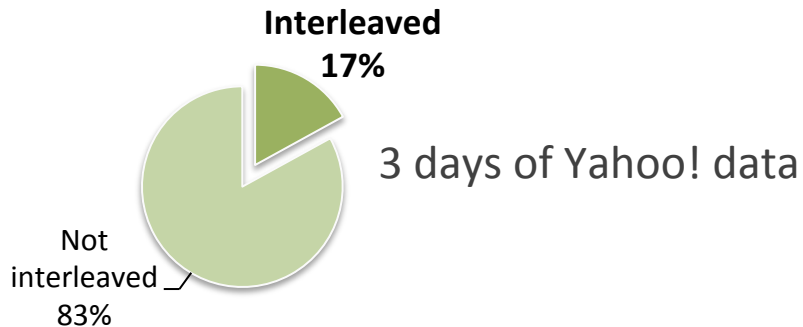
michigan unemployed

Likelihood of x-length tasks



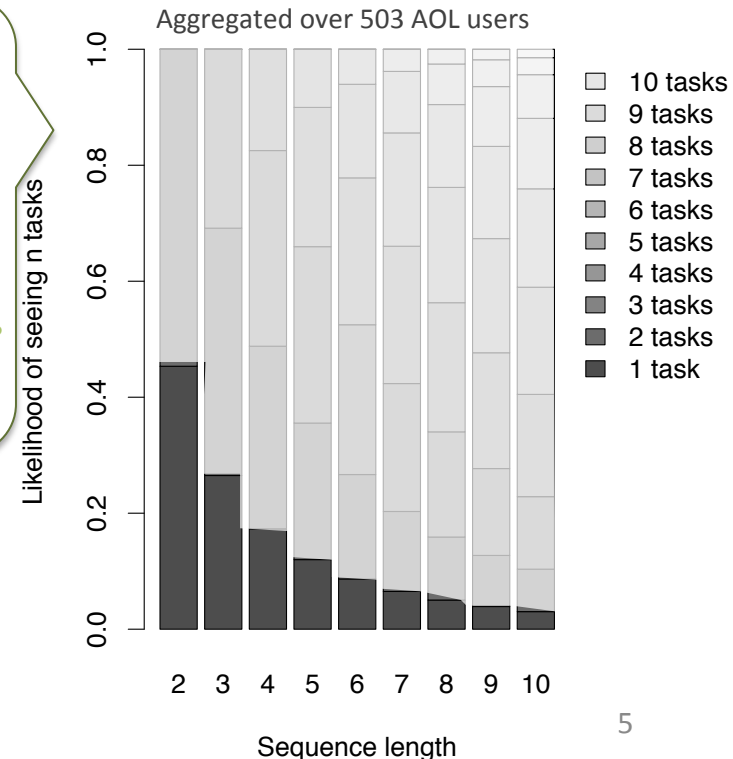
A couple of issues

Interleaved tasks



You are more likely than not to encounter off-task queries even going back *one* query!

Multi-tasking within a fixed window



Ideal system

Identify on-task queries

pampered chef

elliptical trainer

hosting pampered chef

elliptical trainer benefits

Ignore off-task queries

Research questions

1 Does on-task context help?

pampered chef

elliptical trainer

hosting pampered chef

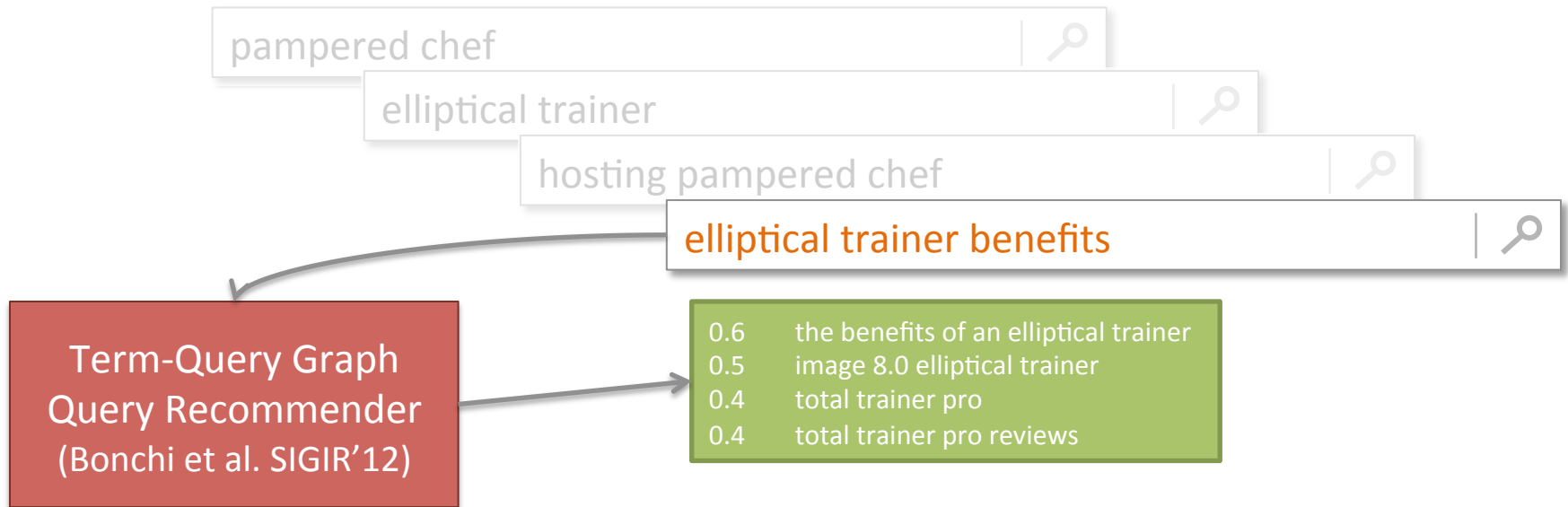
elliptical trainer benefits

2 How much does off-task context hurt?

3 How well does current technology deal with mixed contexts?

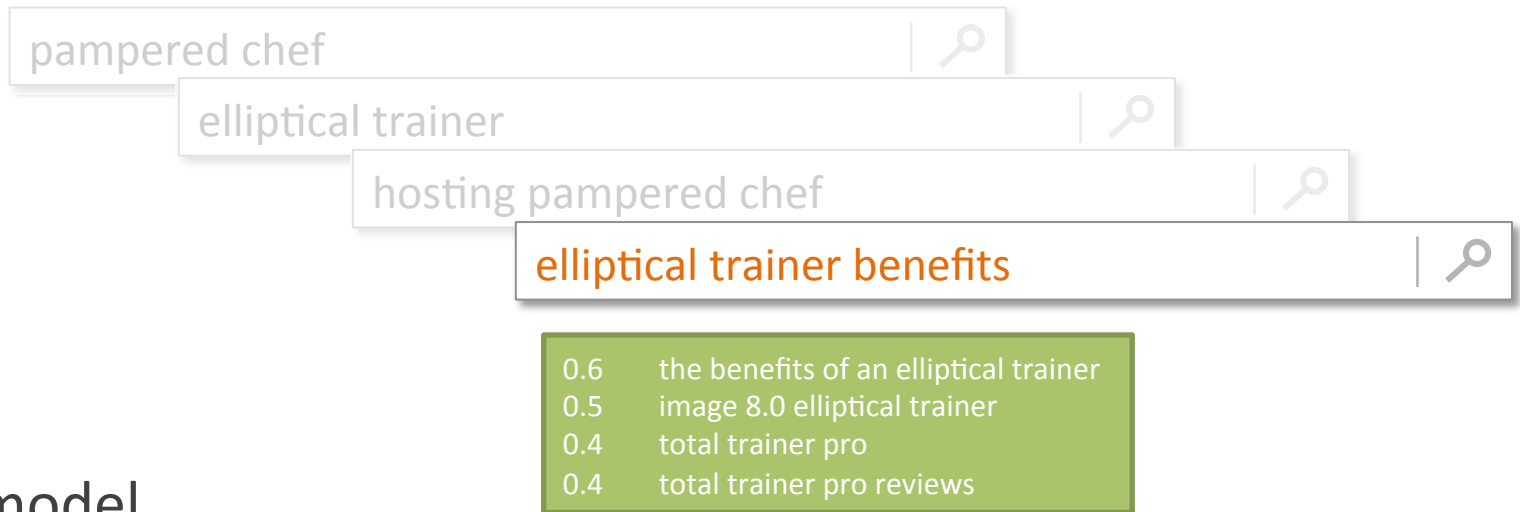
Basic models

Reference query only

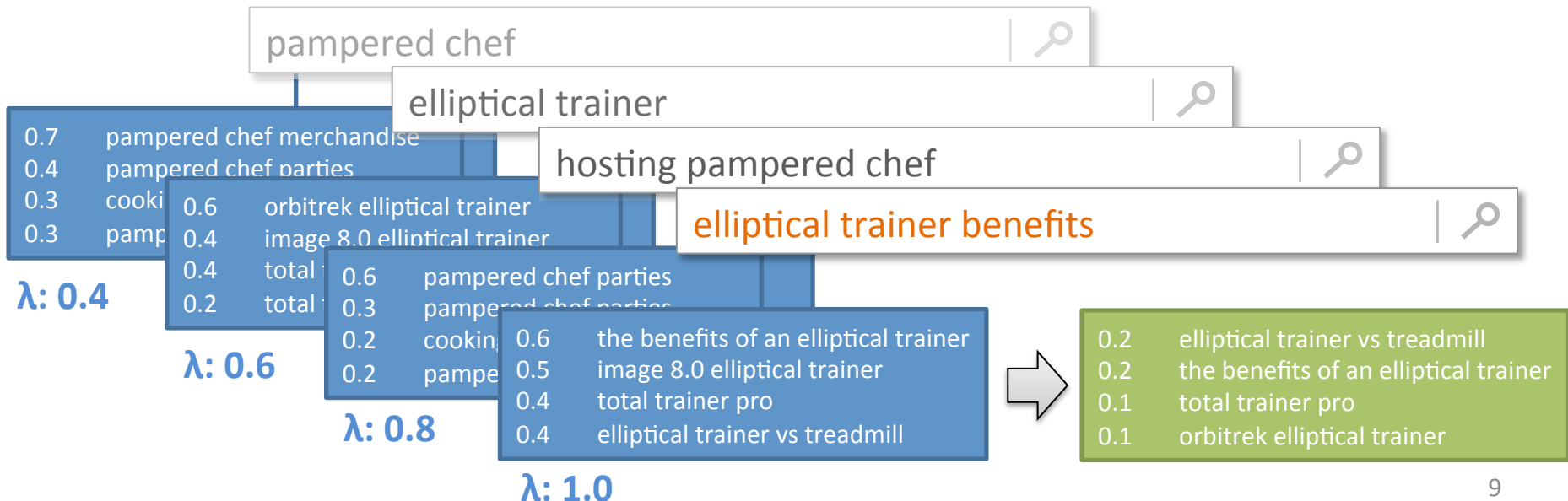


Basic models

Reference query only

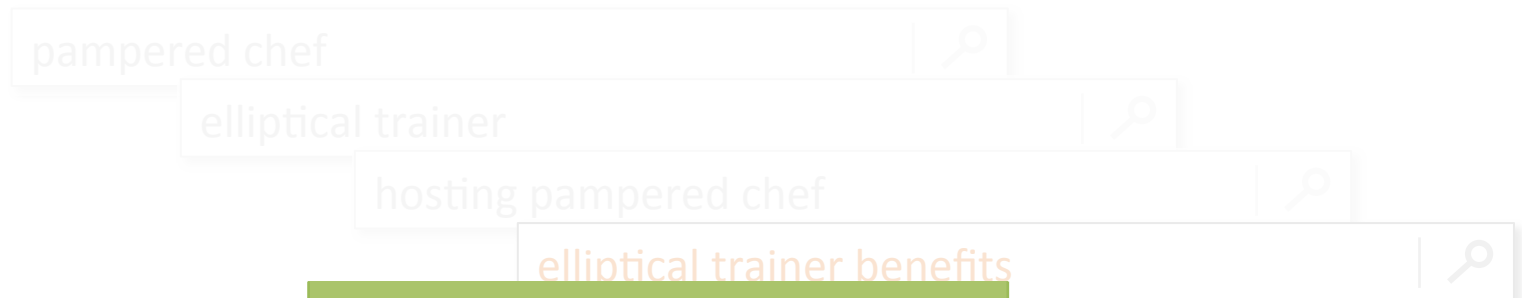


Decay model



Basic models

Reference query only



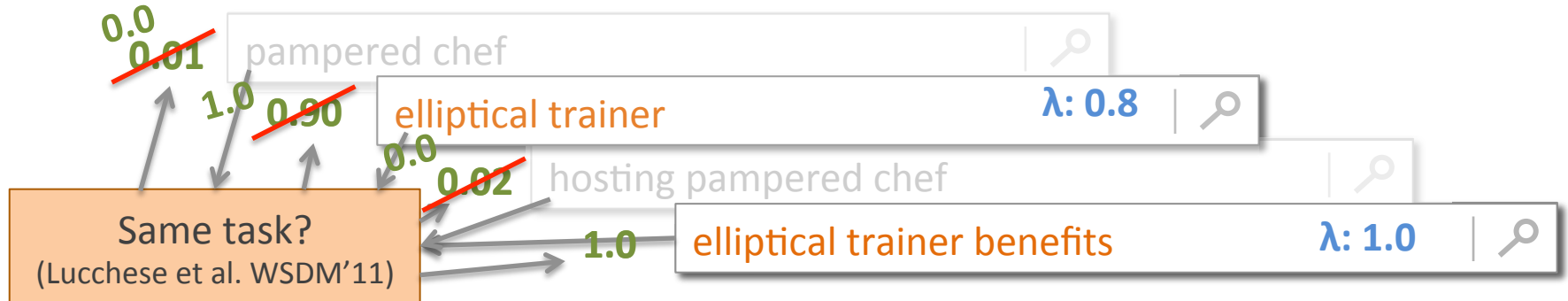
Model	Uses context?
Reference query only	
Decay	✓

Decay model



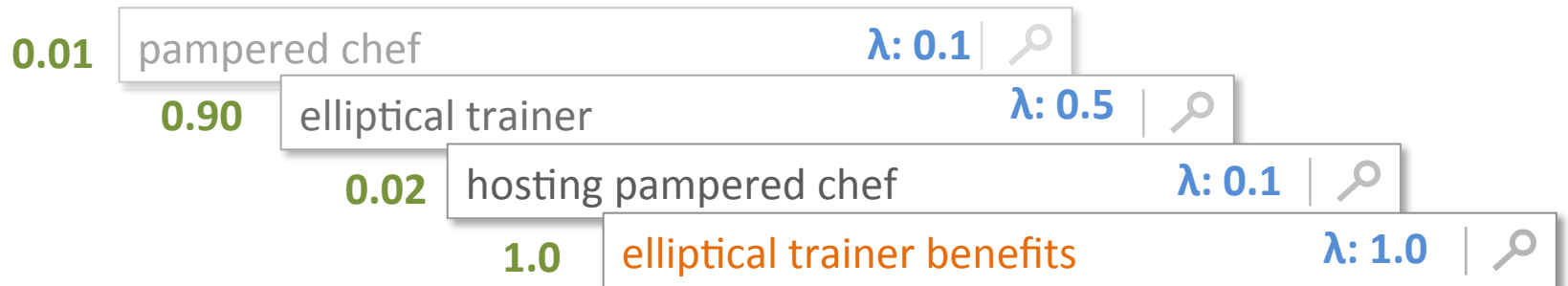
Task-aware models

Hard task threshold model



$\text{weight} = \text{task_decay}$ (decay over on-task queries only)

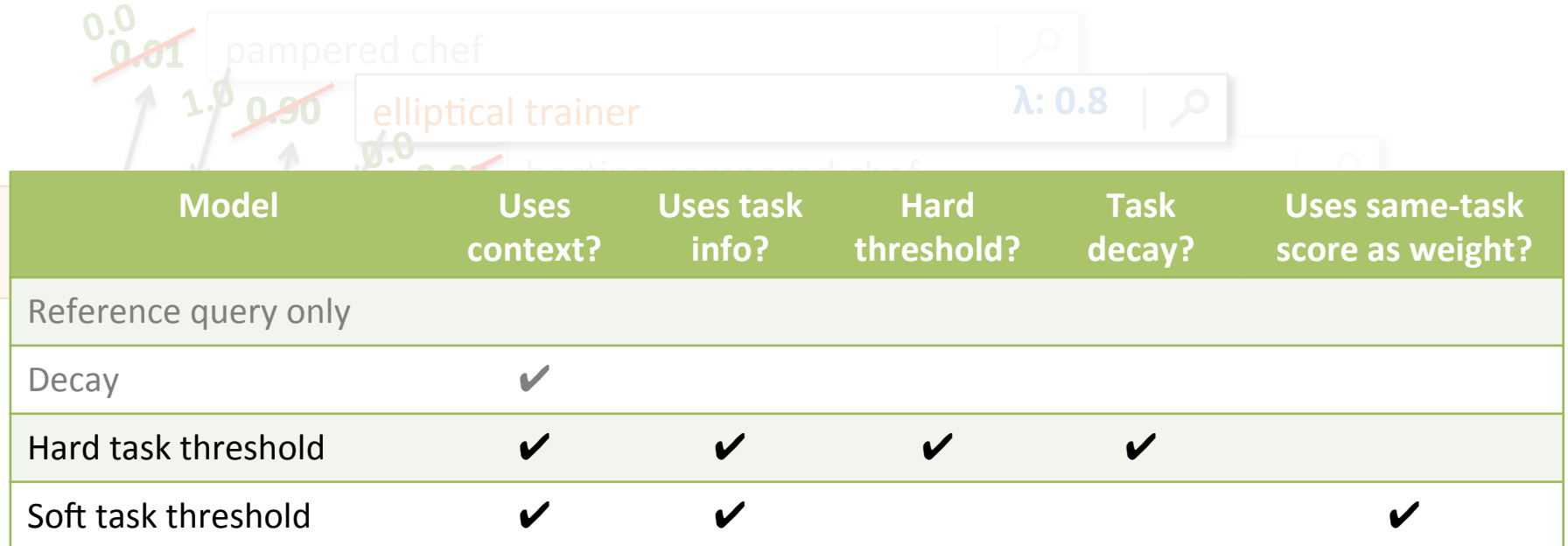
Soft task threshold model



$\text{weight} = \text{decay} \times \text{same_task_score}$

Task-aware models

Hard task threshold model



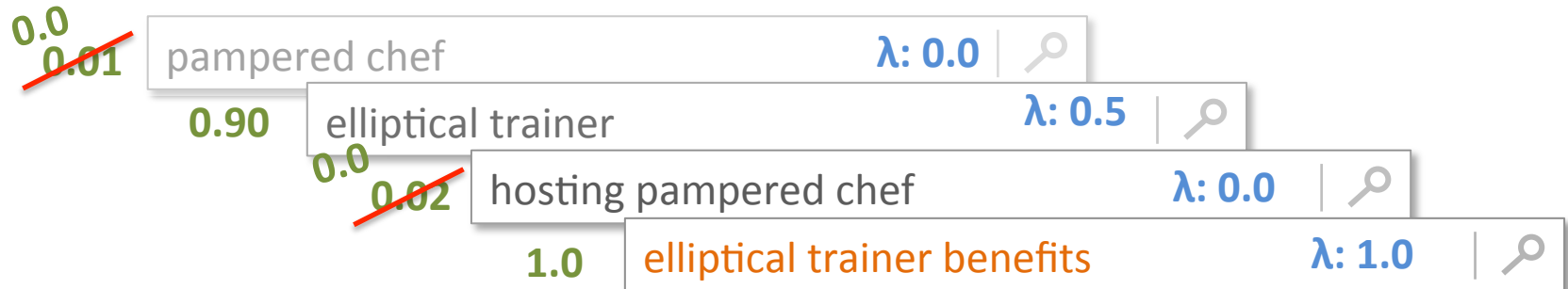
The background image shows a series of search results for the query 'pampered chef'. The results are ranked by a weight, with 'elliptical trainer' being the top result. The weights are shown as λ values: 0.8 for 'elliptical trainer', 0.1 for 'pampered chef', 0.5 for 'elliptical trainer', 0.1 for 'hosting pampered chef', and 1.0 for 'elliptical trainer benefits'. The scores for the top results are 0.01, 0.90, 0.02, and 1.0 respectively. The scores are shown in green, and the weights are shown in blue.

Model	Uses context?	Uses task info?	Hard threshold?	Task decay?	Uses same-task score as weight?
Reference query only					
Decay	✓				
Hard task threshold	✓	✓	✓	✓	
Soft task threshold	✓	✓			✓

$$\text{weight} = \text{decay} \times \text{same_task_score}$$

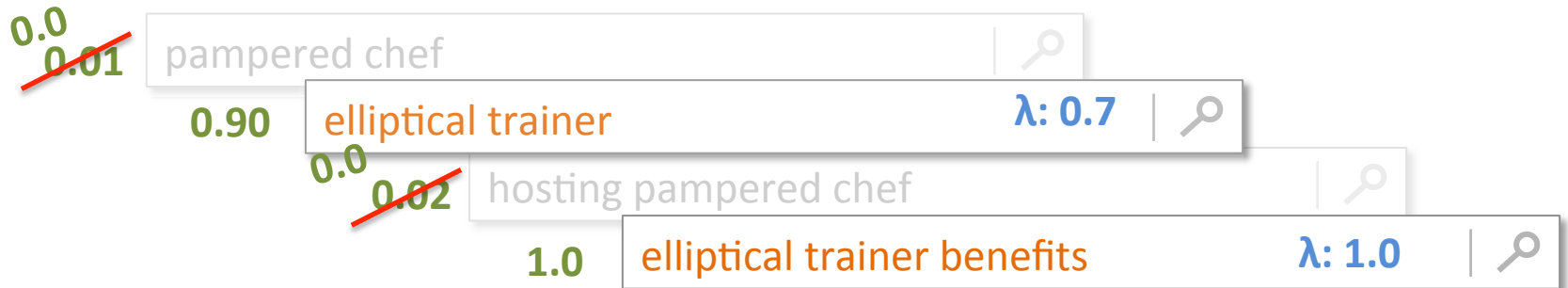
Task-aware models (cont'd)

Firm task threshold model 1



$weight = \text{decay} \times \text{same_task_score}$ if on-task; 0 otherwise
(soft task score for on-task queries; 0 for off-task queries)

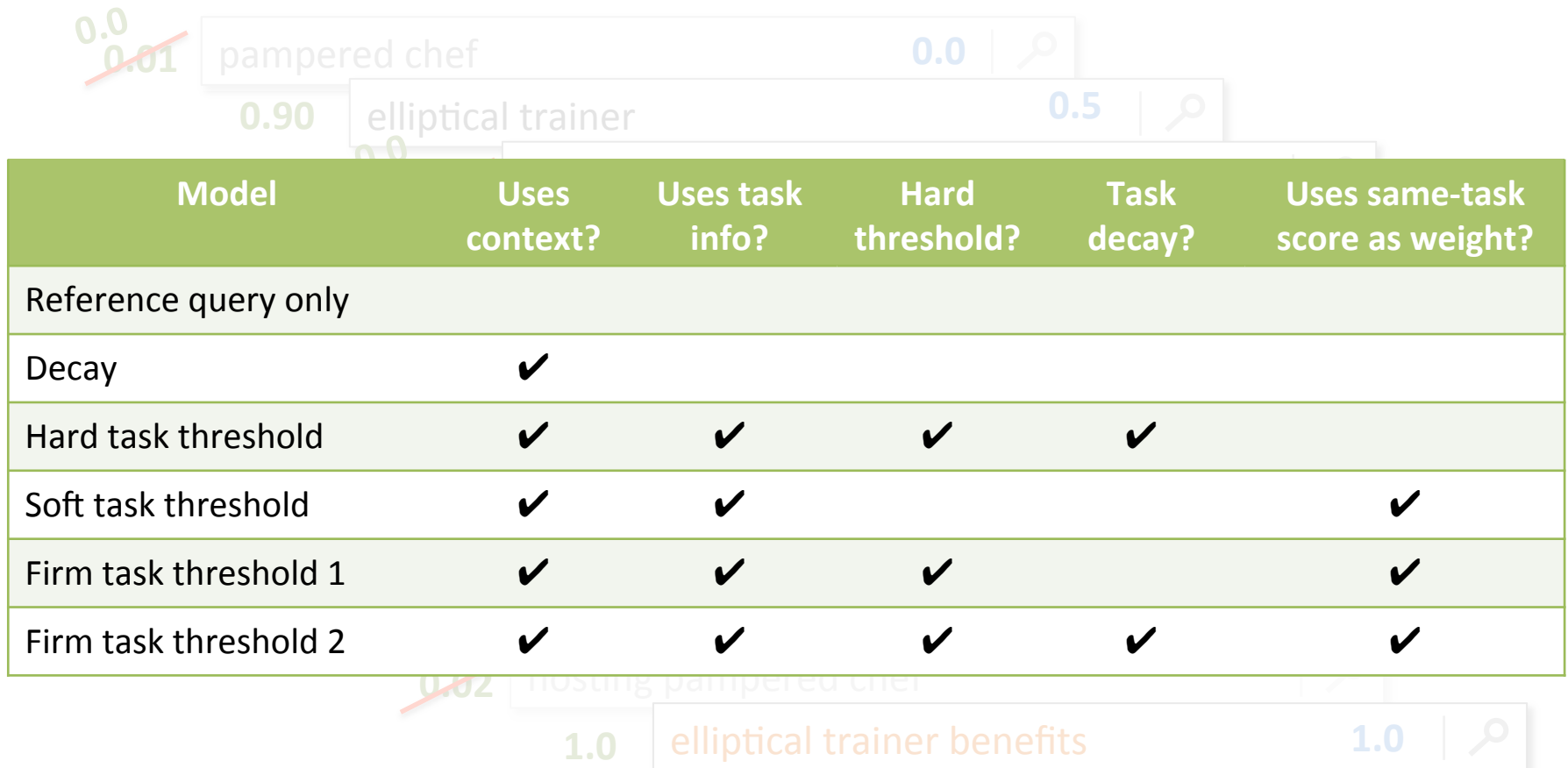
Firm task threshold model 2



$weight = \text{task_decay} \times \text{same_task_score}$

Task-aware models (cont'd)

Firm task threshold model 1

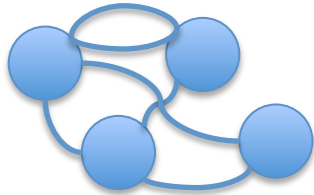


Model	Uses context?	Uses task info?	Hard threshold?	Task decay?	Uses same-task score as weight?
Reference query only					
Decay	✓				
Hard task threshold	✓	✓	✓	✓	
Soft task threshold	✓	✓			✓
Firm task threshold 1	✓	✓	✓		✓
Firm task threshold 2	✓	✓	✓	✓	✓

$$\text{weight} = \text{task_decay} \times \text{same_task_score}$$

Data and evaluation

Query recommendations: 2006 AOL search log



used to build a Term-Query Graph

Tasks: 2010—2011 TREC Session Track

- 212 judged sessions
- 2+ queries/session
- task = session

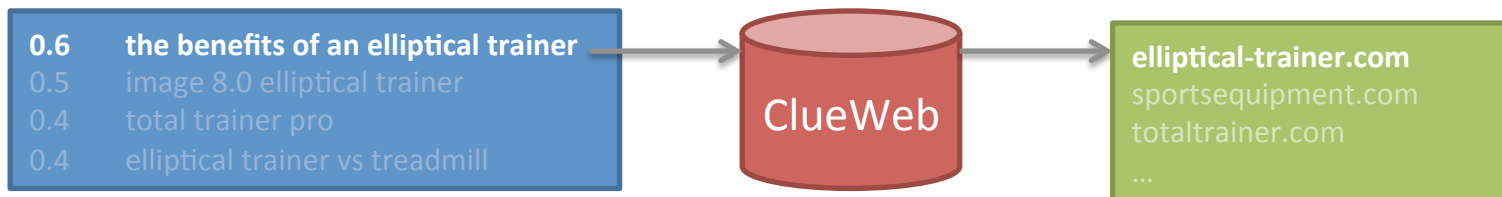
On-task contexts

Off-task contexts

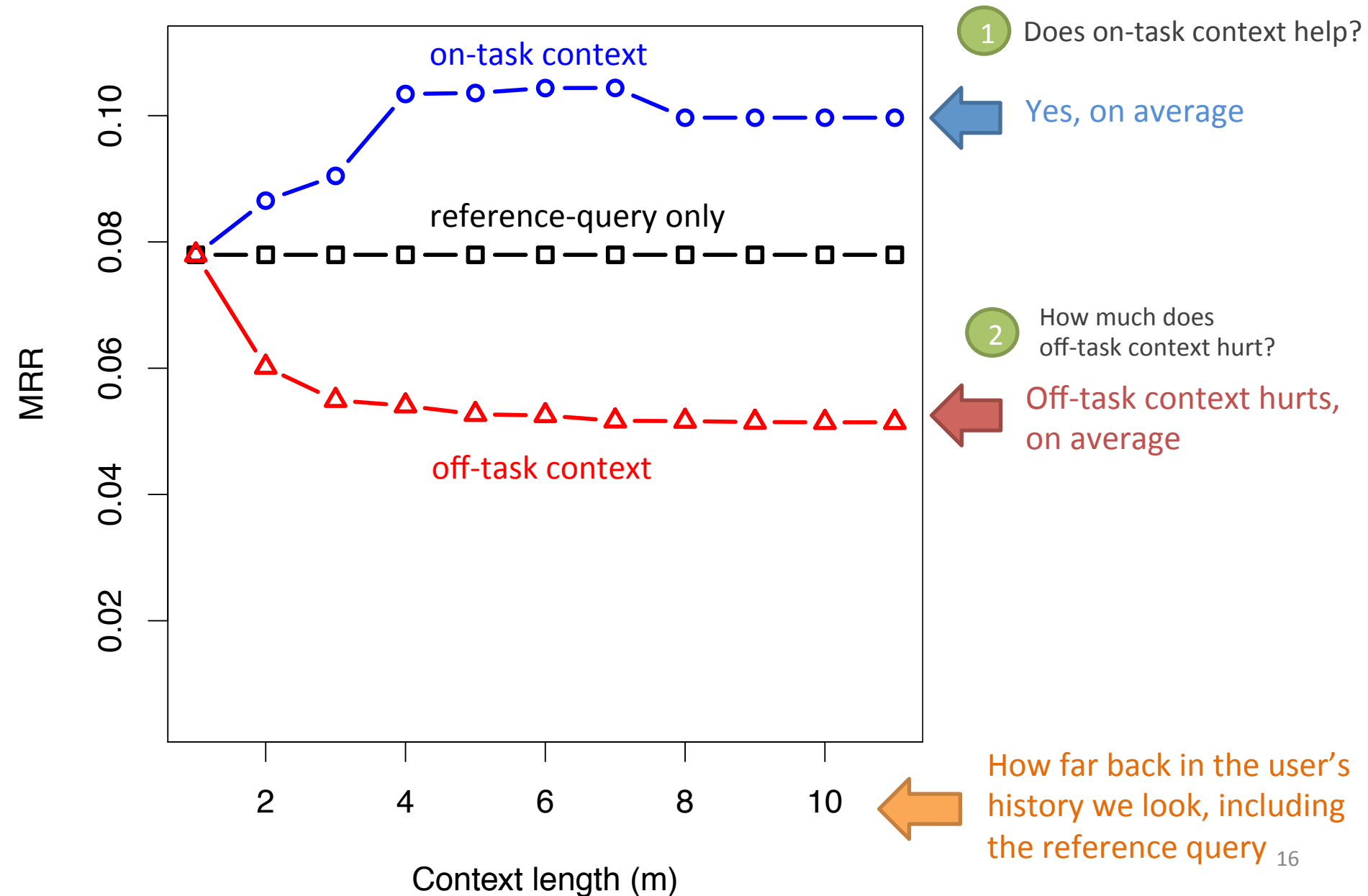
Mixed contexts

Evaluation

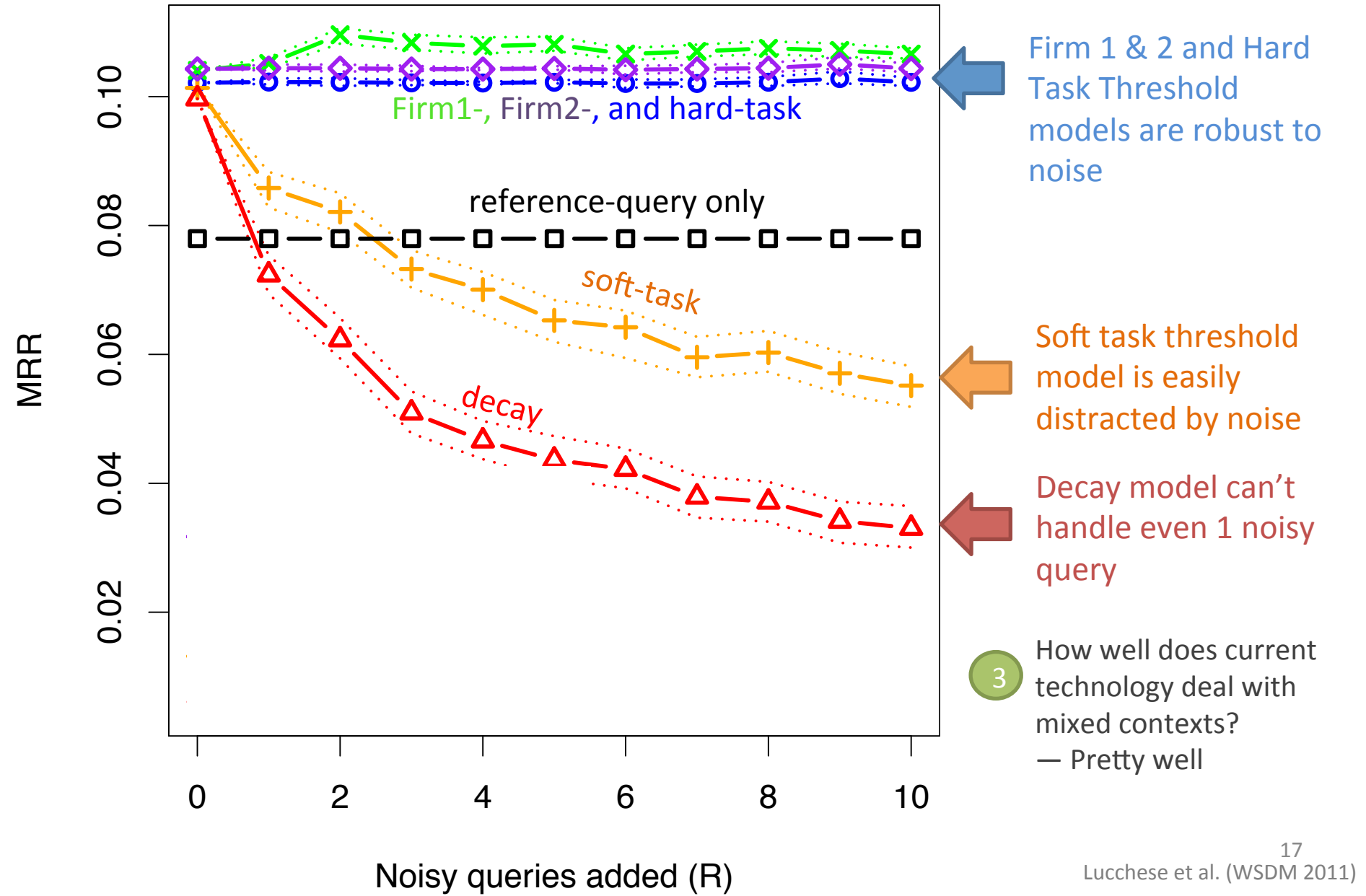
- **When's the soonest a user can find a novel document?**
- MRR over documents retrieved with the top scoring recommendation (ClueWeb '09)
- removed documents retrieved in top 10 of context queries



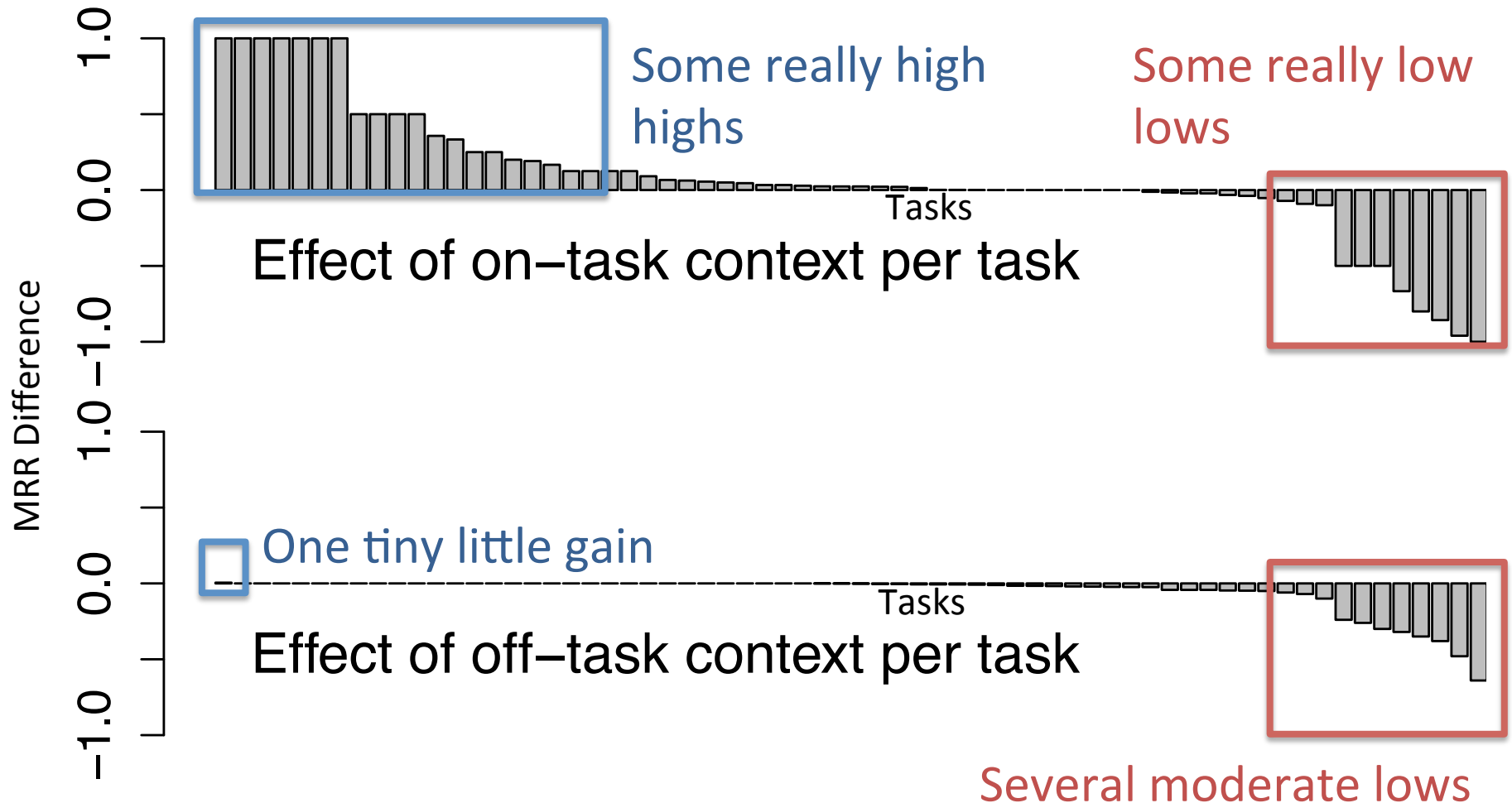
Effect of on-/off-task context



Effect of noise (mixed contexts)

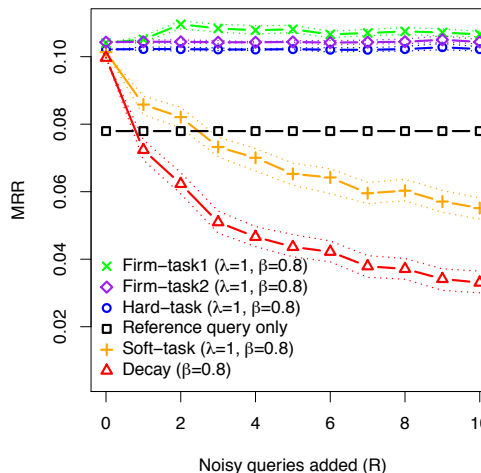
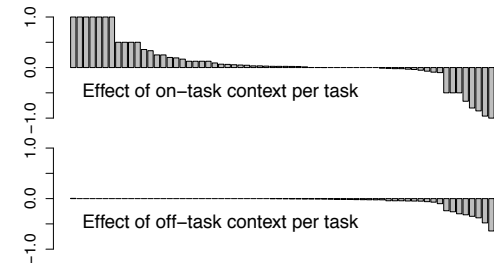


Variance across task set



Conclusions

- on-task context can be very helpful
 - it can also hurt
- off-task context is bad
- current state-of-the-art task segmentation works well



Future work

- do these results hold for other query recommendation algorithms?
- are our findings consistent across additional tasks?
- can we predict when context will help?

Thanks!

Example

Context:

- 5. (1.00) satellite internet providers
- 4. (0.44) hughes internet
- 3. (0.02) ocd beckham
- 2. (0.04) reviews xc90
- 1. (0.04) buy volvo semi trucks



same-task score

Reference-query only recommendations:

	0.08	alabama satellite internet providers
	0.25	sattelite internet
MRR	1.00	satellite internet providers
	0.02	satelite internet for 30.00
	0.06	satellite internet providers northern california

Decay recommendations:

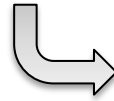
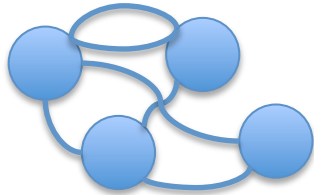
	0.00	2006 volvo xc90 reviews
	0.00	volvo xc90 reviews
MRR	1.00	hughes internet.com
	1.00	hughes satelite internet
	0.08	alabama satellite internet providers

Hard task recommendations:

	1.00	hughes internet.com
	1.00	hughes satelite internet
MRR	0.08	alabama satellite internet providers
	0.17	satellite high speed internet
	0.25	sattelite internet

Data and evaluation

Query recommendations: 2006 AOL search log



used to build a Term-Query Graph

Tasks: 2010—2011 TREC Session Track

On-task contexts

- 212 judged sessions
- 2+ queries/session
- task = session

Off-task contexts

- 50×212 sessions
- all queries but the last are replaced by queries from other tasks
- 50 samples per TREC session

Mixed contexts

- $10 \times 50 \times 212$ sessions
- added up to 10 off-task queries per session
- 50 samples per TREC session and noise level

Evaluation

- MRR over documents retrieved with the top scoring recommendation (ClueWeb '09)
- removed documents retrieved in top 10 of context queries