

# Information Retrieval

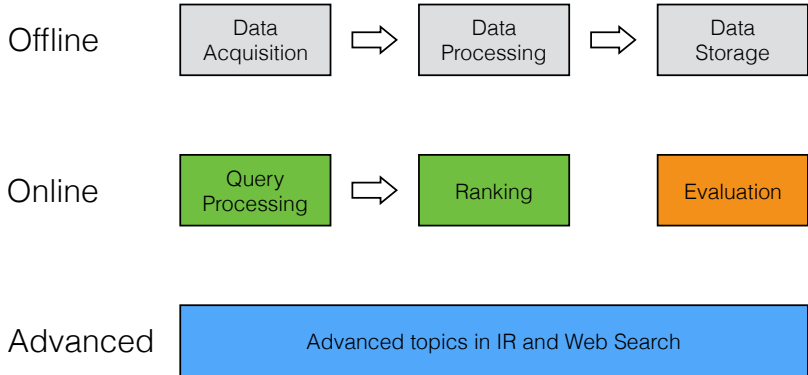
## Click Models

**Ilya Markov**

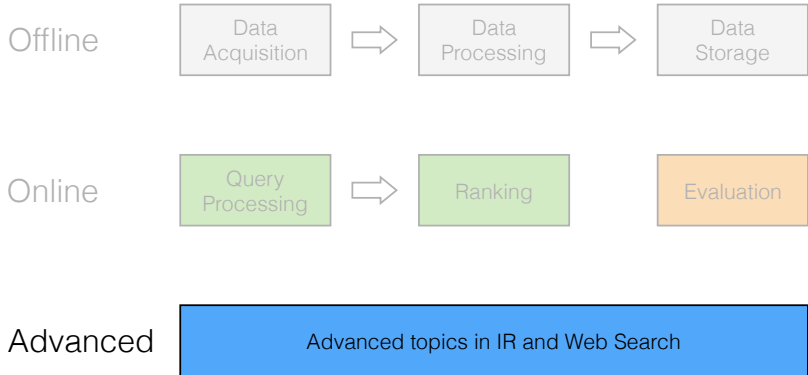
i.markov@uva.nl

University of Amsterdam

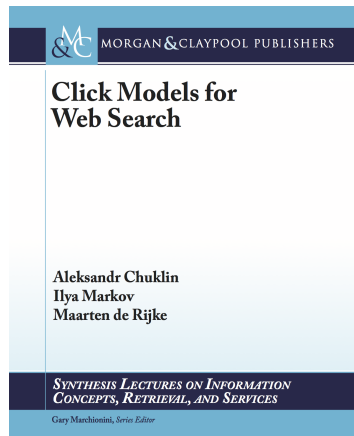
# Course overview



# Advanced topics in IR



# The book



<http://clickmodels.weebly.com/the-book.html>

# Tutorials

- SIGIR 2015, Santiago, Chile
- AINL-ISMW FRUCT 2015, St. Petersburg, Russia
- WSDM 2016, San Francisco, USA
- RuSSIR 2016, Saratov, Russia

<http://clickmodels.weebly.com/tutorials.html>

# Outline

- 1 Current developments
- 2 Future research
- 3 Summary

# Outline

- 1 Current developments
- 2 Future research
- 3 Summary

# Incorporating clicks, attention and satisfaction

## Incorporating clicks, attention and satisfaction into a search engine result page evaluation model

Aleksandr Chuklin, Maarten de Rijke

*Proceedings of CIKM 2016, Indianapolis, USA*



# A neural click model for web search

## A neural click model for web search

Alexey Borisov, Ilya Markov, Maarten de Rijke, Pavel Serdyukov

*Proceedings of WWW 2016, Montreal, Canada*

# A context-aware time model for web search

## A context-aware time model for web search

Alexey Borisov, Ilya Markov, Maarten de Rijke, Pavel Serdyukov


*Proceedings of SIGIR 2016, Pisa, Italy*  
best student paper award


# Time between user actions


- **Time between clicks**
- Time to first click
- Time to last click
- Time between queries


# Time between clicks


Yandex


Web  [Amsterdam travel guide - Wikitravel](#)  
 wikitravel.com > Amsterdam


Images  Amsterdam, capital of the Netherlands. With more than one million inhabitants in its urban area, it is the country's largest city and its financial, cultural, and creative centre.

Video  Amsterdam gives its name from the city's origin as "Dam" ...


Translate  [Amsterdam - Wikipedia, the free encyclopedia](#)  
 en.wikipedia.org > Amsterdam


More  Amsterdam (ˌæmstərˈdæm; Dutch: [ɑmstərˈdɑm]) is the capital and most populous city in the Kingdom of the Netherlands. Its status as the Dutch capital is mandated by the Constitution of the Netherlands though it is not the seat of the Du...


 [Amsterdam Tourism: Best of Amsterdam, The Netherlands](#)  
 tripadvisor.com > Tourism-g188590-Amsterdam\_North... >  
 Amsterdam Tourism: TripAdvisor has 821,053 reviews of Amsterdam Hotels, Attractions, and Restaurants making it your best Amsterdam resource.


 [Your guide to visit, enjoy, live, work & invest in Amsterdam](#)  
 iamsterdam.com > en >  
 Welcome to Amsterdam.com. We would like to ask a few questions about your experience on our website. It will only take a few minutes of your time.


Yandex


Web  [Amsterdam travel guide - Wikitravel](#)  
 wikitravel.com > Amsterdam


Images  Amsterdam, capital of the Netherlands. With more than one million inhabitants in its urban area, it is the country's largest city and its financial, cultural, and creative centre.

Video  Amsterdam gives its name from the city's origin as "Dam" ...

Translate  [Amsterdam - Wikipedia, the free encyclopedia](#)  
 en.wikipedia.org > Amsterdam

More  Amsterdam (ˌæmstərˈdæm; Dutch: [ɑmstərˈdɑm]) is the capital and most populous city in the Kingdom of the Netherlands. Its status as the Dutch capital is mandated by the Constitution of the Netherlands though it is not the seat of the Du...

 [Amsterdam Tourism: Best of Amsterdam, The Netherlands](#)  
 tripadvisor.com > Tourism-g188590-Amsterdam\_North... >  
 Amsterdam Tourism: TripAdvisor has 821,053 reviews of Amsterdam Hotels, Attractions, and Restaurants making it your best Amsterdam resource.

 [Your guide to visit, enjoy, live, work & invest in Amsterdam](#)  
 iamsterdam.com > en >  
 Welcome to Amsterdam.com. We would like to ask a few questions about your experience on our website. It will only take a few minutes of your time.

} 30 secs

there is *context bias*

# Modeling time

- Average

$$Time("Amsterdam", "wikipedia.org") = \frac{120 + 60 + 30}{3}$$

- Probability distribution

$$Time("Amsterdam", "wikipedia.org") \sim \text{Gamma}(\mathbf{k}, \boldsymbol{\theta})$$

where  $(\mathbf{k}, \boldsymbol{\theta})$  are estimated from 120, 60, 30

context bias is not modeled

# Context-aware time modeling

$Time(\text{"Amsterdam"}, \text{"wikipedia.org"}, context_1) \sim \text{Gamma}(\mathbf{k}_1, \theta_1)$

$Time(\text{"Amsterdam"}, \text{"wikipedia.org"}, context_2) \sim \text{Gamma}(\mathbf{k}_2, \theta_2)$

$Time(\underbrace{\text{"Amsterdam"}, \text{"wikipedia.org"}}_{\text{user action}}, \underbrace{context_3}_{\text{context}}) \sim \text{Gamma}(\mathbf{k}_3, \theta_3)$

# Context-aware time modeling

$$\begin{aligned} \text{Time}(\text{action}, \text{context}) \sim \text{Gamma}(\quad \\ \mathbf{a}_k(\text{ctx}) \cdot \mathbf{k}(\text{act}) + \mathbf{b}_k(\text{ctx}), \\ \mathbf{a}_\theta(\text{ctx}) \cdot \boldsymbol{\theta}(\text{act}) + \mathbf{b}_\theta(\text{ctx})) \end{aligned}$$

# Parameter estimation

$$\begin{aligned} \text{Time}(\text{action}, \text{context}) \sim \text{Gamma}(\quad \\ \mathbf{a}_k(\text{ctx}) \cdot \mathbf{k}(\text{act}) + \mathbf{b}_k(\text{ctx}), \\ \mathbf{a}_\theta(\text{ctx}) \cdot \boldsymbol{\theta}(\text{act}) + \mathbf{b}_\theta(\text{ctx})) \end{aligned}$$

- 1 Fix **context-independent** parameters
- 2 Optimize **context-dependent** parameters using *neural networks*
- 3 Fix **context-dependent** parameters
- 4 Optimize **context-independent** using *gradient descent*
- 5 Repeat until convergence



# Parameter estimation

- We do not know the form of **context-dependent** parameters  
⇒ neural networks
- We know the form of **context-independent** parameters  
(Gamma distribution) ⇒ direct optimization

# Context

| General  |                 |
|--|-----------------|
| Is query (Q-action)                                    | (0: no, 1: yes) |
| Is click (C-action)                                    | (0: no, 1: yes) |
| $\log(1 + \text{observed time since previous action})$ | (0: undefined)  |
| $\log(1 + \text{average time since previous action})$  | (0: undefined)  |
| Q-action   |                 |
| Is new search session                                  | (0: no, 1: yes) |
| Number of terms in issued query                        | (0: undefined)  |
| BM25 (issued query, previous query)                    | (0: undefined)  |
| BM25 (previous query, issued query)                    | (0: undefined)  |
| C-action   |                 |
| Is click on the 1 <sup>st</sup> position               | (0: no, 1: yes) |
| ...  | ...             |
| Is click on the 10 <sup>th</sup> position              | (0: no, 1: yes) |

# Dataset

3 months of log data from Yandex search engine

| Time between actions      | Max time | # Observations |
|---------------------------|----------|----------------|
| Time-to-first-click       | 1 min    | 30,747,733     |
| Time-between-clicks       | 5 min    | 6,317,834      |
| Time-to-last-click        | 5 min    | 30,446,973     |
| Time-from-abandoned-query | 1 min    | 11,523,351     |

# Evaluation tasks

**Task1.** Predict time between clicks

- Log-likelihood
- Root mean squared error (MSE)

**Task2.** Rank results based on time between clicks

- $n\text{DCG}@\{1, 3, 5, 10\}$

# Task 1. Predicting time

| Time model    | Distribution | Log-likelihood | RMSE  |
|---------------|--------------|----------------|-------|
| Basic         | exponential  | -4.9219        | 60.73 |
|               | gamma        | -4.9105        | 60.76 |
|               | Weibull      | -4.9077        | 60.76 |
| Context-aware | exponential  | -4.8787        | 58.93 |
|               | gamma        | -4.8556        | 58.98 |
|               | Weibull      | -4.8504        | 58.94 |

## Task 2. Ranking results

| Time model    | Distribution | NDCG  |       |       |       |
|---------------|--------------|-------|-------|-------|-------|
|               |              | @1    | @3    | @5    | @10   |
| Average       | —            | 0.651 | 0.693 | 0.728 | 0.812 |
| Context-aware | exponential  | 0.668 | 0.710 | 0.743 | 0.820 |
|               | gamma        | 0.675 | 0.715 | 0.748 | 0.822 |
|               | Weibull      | 0.671 | 0.709 | 0.745 | 0.821 |

# Other times

- Time to first click
- Time to last click
- Time between queries

# Summary

- Removed **context bias** from time between actions
- Predicted user search interactions better (**Task 1**)
- Used the **context-independent** component for better document ranking (**Task 2**)



# Outline

- 1 Current developments
- 2 Future research
- 3 Summary

# Future research

- Keep on adding new variables – not a good idea
- Parameter estimation
  - Efficiency
  - Online learning
- Other interactions and environments
  - Interactions beyond clicks
  - Devices beyond desktop computers

# Future research

| Model's output          | Evaluation         | Application         |
|-------------------------|--------------------|---------------------|
| Conditional click probs | Log-likelihood     | User simulation     |
| Full click probs        | Perplexity         | Model-based metrics |
| Parameter values        | Ranking evaluation | Ranking             |

- Why use intermediate evaluation?
  - Evaluate applications, not models
- Why maximize log-likelihood?
  - Optimize models for specific applications

# Outline

- 1 Current developments
- 2 Future research
- 3 Summary**

# Outline

- 1 Current developments
- 2 Future research
- 3 Summary

# Materials

- Aleksandr Chuklin, Ilya Markov, Maarten and de Rijke  
**Click Models for Web Search**  
Morgan & Claypool, 2015

# Advanced topics in IR

