Removing Bias from Search Interaction Data

Ilya Markov
UvA, i.markov@uva.nl

Joint work with Alexey Borisov,
Maarten de Rijke, Pavel Serdyukov
Bias in Search Interactions
Clicks
Dealing with Bias

1. Design tools that help users to overcome their biases (e.g., Search Results Navigator)

2. Learn to remove biases from user interaction data
Removing Bias

\[ P(\text{click}) = P(\text{look}) \cdot P(\text{like}) \]

- Result rank
- Presence of images
- Previous clicks
- Etc.
Click Models for Web Search

Aleksandr Chuklin
Ilya Markov
Maarten de Rijke

SYNTHESIS LECTURES ON INFORMATION CONCEPTS, RETRIEVAL, AND SERVICES

Gary Marchionini, Series Editor
Time between Actions
Dwell Time

Amsterdam - Wikipedia, the free encyclopedia
en.wikipedia.org > Amsterdam
Amsterdam (Amsterdam, [ɑmˈstɑrˌdam],; Dutch: [ˈɑmstər(ˌ)dɑm]) 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...
Time between Clicks
Modeling Time

- Average

\[
\text{Time}(\text{“Amsterdam”, “wikipedia.org”}) = \frac{120 + 60 + 30}{3}
\]

- Probability distribution

\[
\text{Time}(\text{“Amsterdam”, “wikipedia.org”}) \sim \text{Gamma}(k, \theta)
\]

\((k, \theta)\) are estimated from 120, 60, 30
Context-aware Modeling

\[
\text{Time(“Amsterdam”, “wikipedia.org”, context}_1) \sim \Gamma(k_1, \theta_1)
\]

\[
\text{Time(“Amsterdam”, “wikipedia.org”, context}_2) \sim \Gamma(k_2, \theta_2)
\]

\[
\text{Time(“Amsterdam”, “wikipedia.org”, context}_3) \sim \Gamma(k_3, \theta_3)
\]
Context-aware Modeling

\[
\text{Time}(\text{action}, \text{context}) \sim \Gamma(a_k(ctx), b_k(ctx)) + \Gamma(a_\theta(ctx), b_\theta(ctx))
\]

context-dependent

context-independent
Learning Parameters

\[ Time(\text{action}, \text{context}) \sim \text{Gamma}( \]
\[ a_k(\text{ctx}) \cdot k(\text{act}) + b_k(\text{ctx}), \]
\[ a_\theta(\text{ctx}) \cdot \theta(\text{act}) + b_\theta(\text{ctx}) \]
\]

- Fix one type of parameters, update the other type, repeat
- Neural networks for context-dependent params
- Direct optimization for context-independent params
Evaluation

• Three months of logs from Yandex (6+ million time observations)

• **Task 1.** Predict time between clicks

• **Task 2.** Rank results based on time between clicks
Task 1. Predicting Time

\[ \text{Gamma}(a_k \cdot k + b_k, a_\theta \cdot \theta + b_\theta) \]
### Task 1. Predicting Time

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic</td>
<td>— 4.92</td>
<td>60.73</td>
</tr>
<tr>
<td>context-aware</td>
<td>— 4.88</td>
<td>58.93</td>
</tr>
</tbody>
</table>
Task 2. Ranking Results

Yandex

amsterdam — 191 million answers

Amsterdam travel guide - Wikitruvel

Amsterdam is the 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. Amsterdam derives its name from the city's origin as "Dam" ...

average time

Your guide to visit, enjoy, live, work & invest in Amsterdam

Welcome to I amsterdam.com. We would like to ask a few questions about your experience on our website. This will only take a few minutes of your time.

average time

Amsterdam - Wikipedia, the free encyclopedia

Amsterdam (/ˈæmstərˌdæm, ˈæmstər ˈdæm/; Dutch: [ˈɑmstɛrˌdɑm]) is the capital and most populous city of 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...

average time

Amsterdam Tourism: Best of Amsterdam, The Netherlands

Amsterdam Tourism: TripAdvisor has 821,053 reviews of Amsterdam Hotels, Attractions, and Restaurants making it your best Amsterdam resource.

average time
Task 2. Ranking Results

Yandex

amsterdam — 191 million answers

Web
Images
Video
Translate

More

Amsterdam travel guide - Wikitravel
wikitravel.org > Amsterdam
Amsterdam is the 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. Amsterdam derives its name from the city’s origin as “Dam” ...

\( \Gamma(k, \theta) \)

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

\( \Gamma(k, \theta) \)

Amsterdam - Wikipedia, the free encyclopedia
en.wikipedia.org > Amsterdam
Amsterdam (ˈæmstərˌdæm, əˈmɛstər dəm; Dutch: [ˈɑmstər dor]) is the capital and most populous city of 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...

\( \Gamma(k, \theta) \)

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.

\( \Gamma(k, \theta) \)
Task 2. Ranking Results

\[ \text{Gamma}(k, \theta) = a_k \cdot k + b_k, \]
\[ \text{Gamma}(k, \theta) = a_\theta \cdot \theta + b_\theta \]

\[ \text{Amsterdam travel guide - Wikitavel} \]
Amsterdam is the 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. Amsterdam derives its name from the city’s origin as “Dam” ...

\[ \text{Gamma}(k, \theta) \]

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

\[ \text{Gamma}(k, \theta) \]

Amsterdam - Wikipedia, the free encyclopedia
en.wikipedia.org > Amsterdam
Amsterdam (ˈaːmstərˌdam, ˌɑmstərˈdɑːm; Dutch: [ˈɑmstər dom]) is the capital and most populous city of 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...

\[ \text{Gamma}(k, \theta) \]
## Task 2. Ranking Results

<table>
<thead>
<tr>
<th></th>
<th>NDCG@1</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic</td>
<td>0.651</td>
<td>0.812</td>
</tr>
<tr>
<td>context-aware</td>
<td>0.668</td>
<td>0.820</td>
</tr>
</tbody>
</table>
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)
Takeaway

• User interaction data may be biased
• Split it into bias-absorbing and bias-independent components
• Let neural networks learn the sources of bias
• Help users to overcome their biases

Ilya Markov, i.markov@uva.nl
https://staff.fnwi.uva.nl/i.markov