Beyond Position Bias: Constructing More Reliable Click models for Web Search Engines

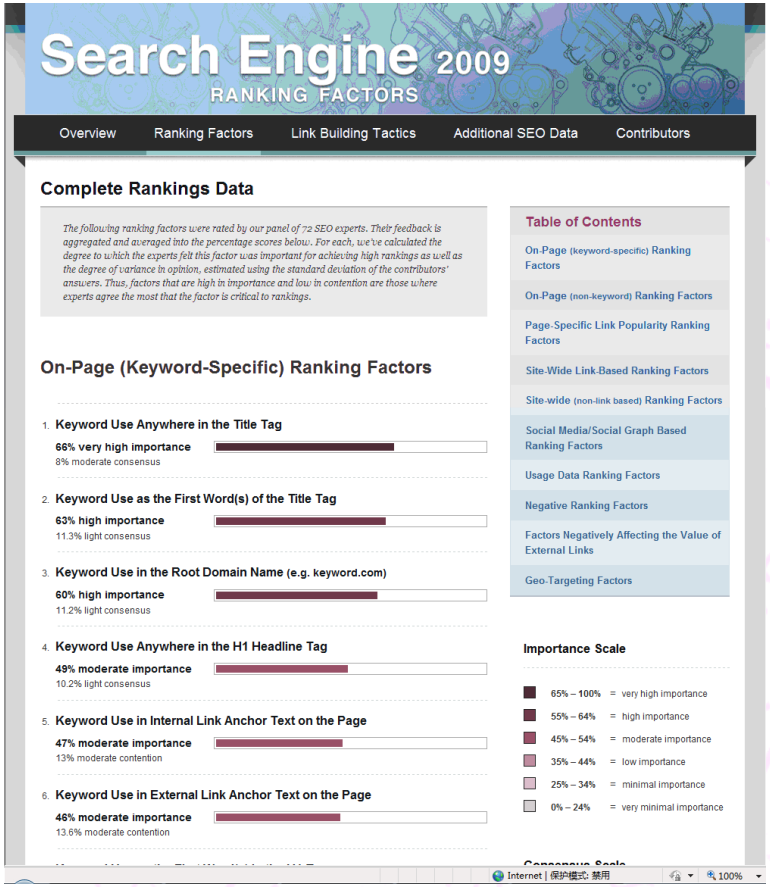
Yiqun Liu

Associate Professor, Tsinghua University

Beijing, China

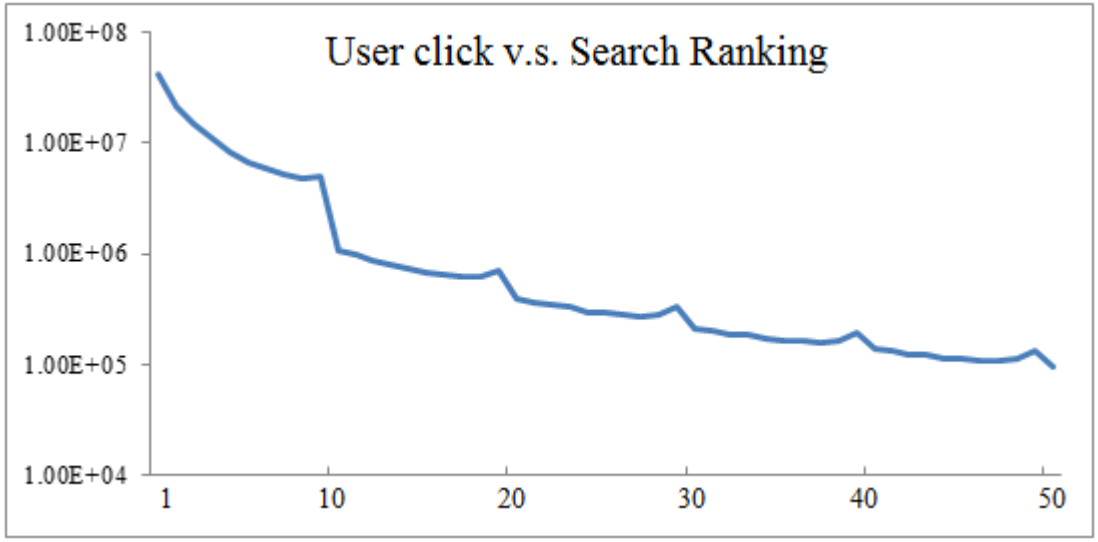
Search Engine Ranking

- How many signals are adopted in search ranking?
  - SEO site: 100+
  - Yahoo LTR task: 700+
  - Content, Hyperlink structure User behavior, Timeliness, Credibility, …
  - Relevance feedback from search users
    - More clicks => higher rankings?



Relevance Feedback

- A naïve idea: user click = voting for relevance
  - 百度 => www.baidu.com; 清华 =>tsinghua.edu.cn
  - 163 => mail.163.com; 搜狗 => d.sogou.com
- Possible problem: position bias



Relevance Feedback

Information Retriever @ Tsinghua University

* Possible problem: presentation bias

* Possible problem: user behavior credibility

Constructing Click Models

* Examination hypothesis to avoid position bias

$$C_i = 1 \rightarrow \quad R_i = 1$$

$$C_i = 1 \rightarrow \quad E_i = 1, R_i = 1$$

  * Cascade model: $P(E_{i+1} = 1 | E_i = 1, C_i) = 1 - C_i$
  * Dependent click model (DCM):

$$P(E_{i+1} = 1 | E_i = 1, C_i = 0) = 1$$

$$P(E_{i+1} = 1 | E_i = 1, C_i = 1) = \lambda_i$$

  * User browsing model (UBM):

$$P(E_i = 1 | C_{1 \ldots i-1}) = \lambda_{r_i, d_i}$$

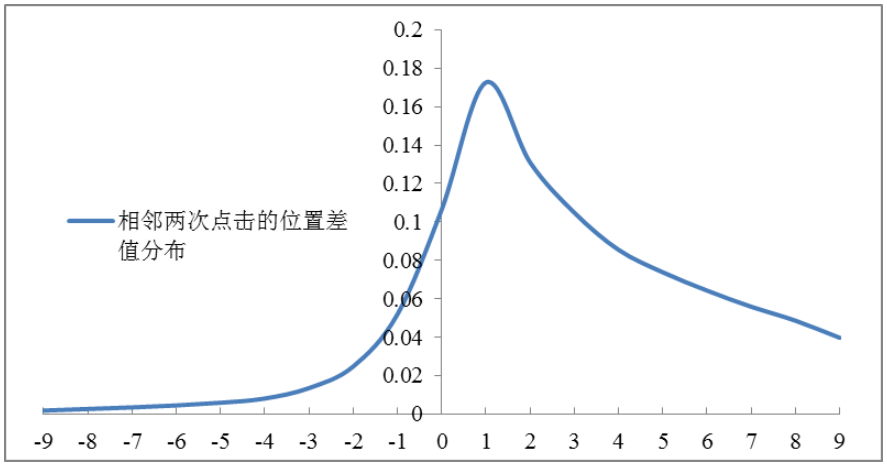
  * Other models: DBM, CCM, ...

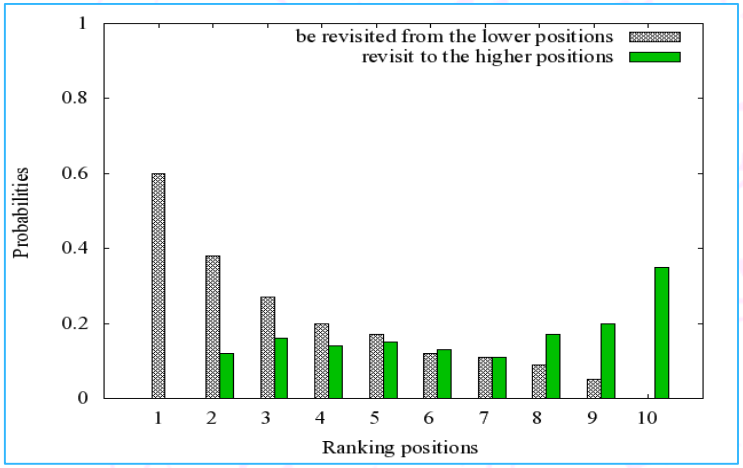
Constructing Click Models

- Problems with existing models
  - Search results are not always examined sequentially
    - Revisit clicks happens a lot
  - Search results do not appear the same
    - Appearance of vertical results are different
  - Users have different behavior preference
    - Some clicks more, some examines more
- Our work: constructing click models considering revisiting / presentation bias / user credibility

Incorporating Revisiting Behaviors

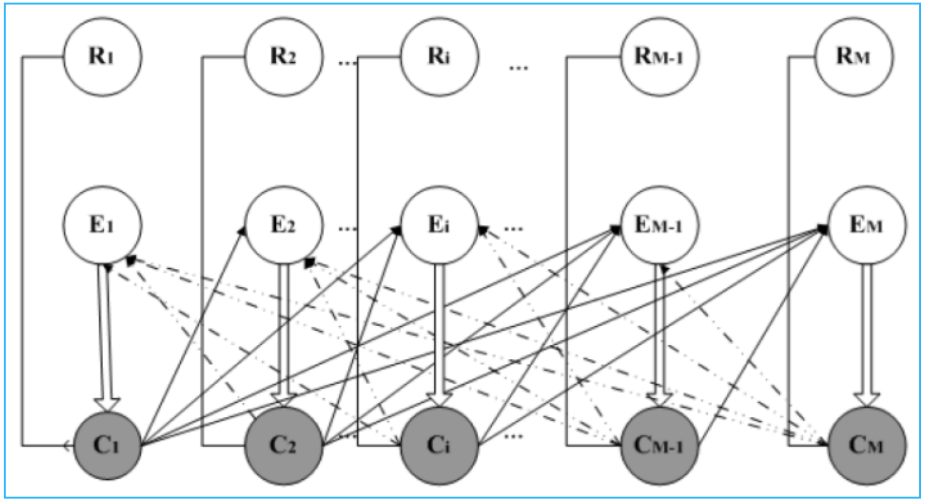
- Revisiting happens a lot for search users
  - Eye tracking experiments (Lorigo et.al, 2005) show that lots of people revisit to previous skipped results
  - Chinese SE (Sogou): 24.1% sessions contain revisiting
  - English SE (Yandex):61.5% sessions contain revisiting





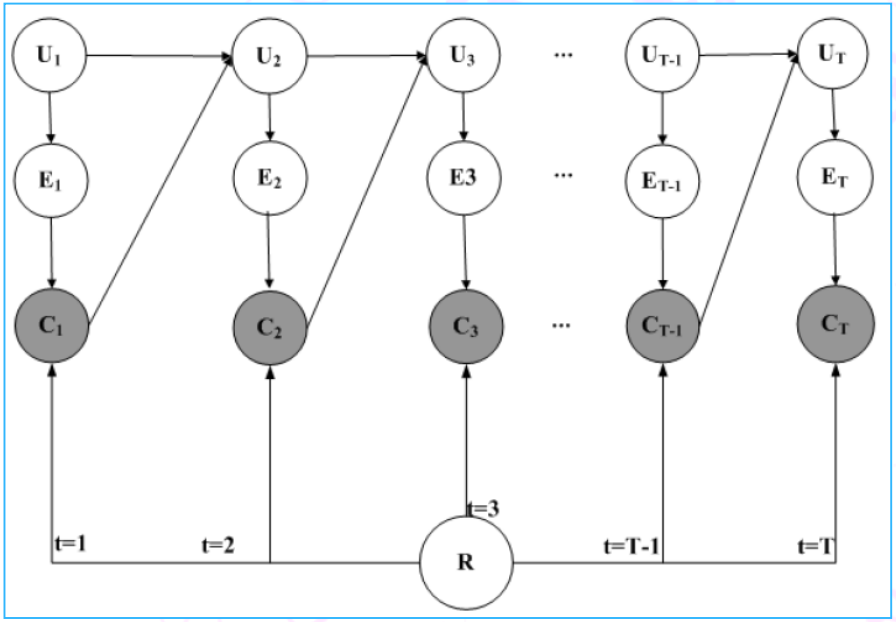
Incorporating Revisiting Behaviors

* THCM: From ranking sequence to time sequence



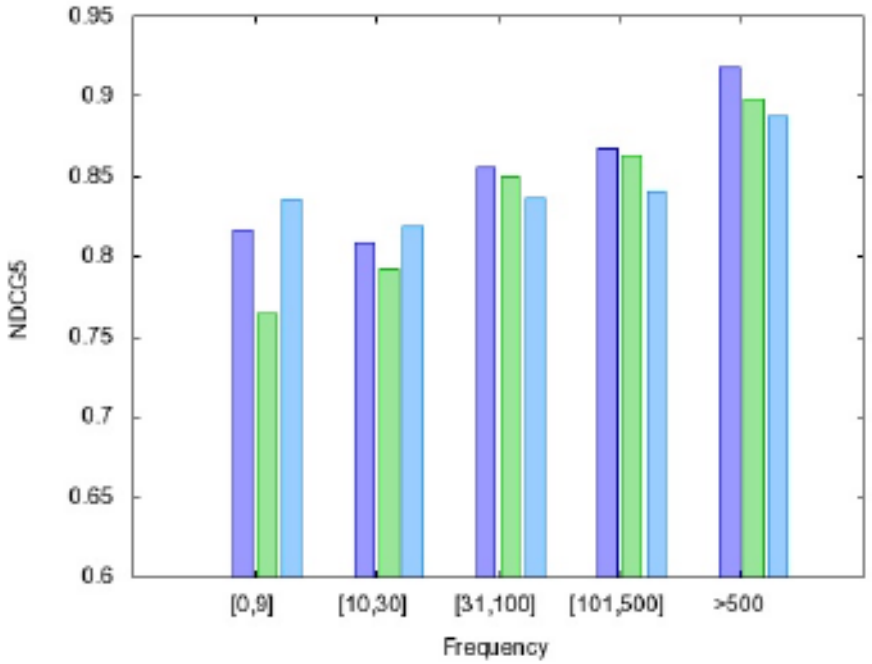
Forward event: $P(E_{t(i+1)}=1 \mid E_{ti}=1)=\alpha$

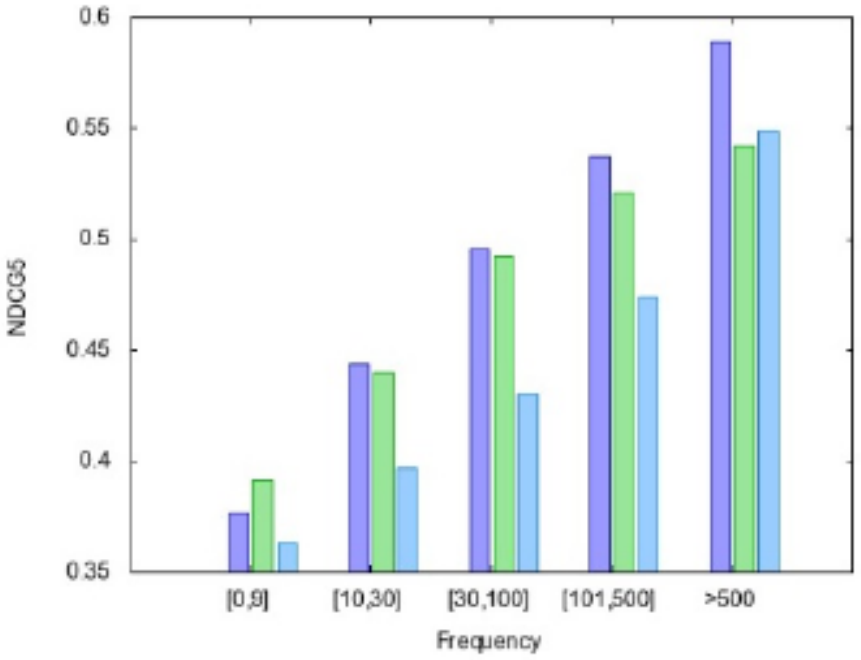
Backward event: $P(E_{t(i-1)}=1 \mid E_{ti}=1)=\gamma$



Incorporating Revisiting Behaviors

* THCM: performance
  * Improvement compared with existing models
  * Works well on both hot and long-tail queries



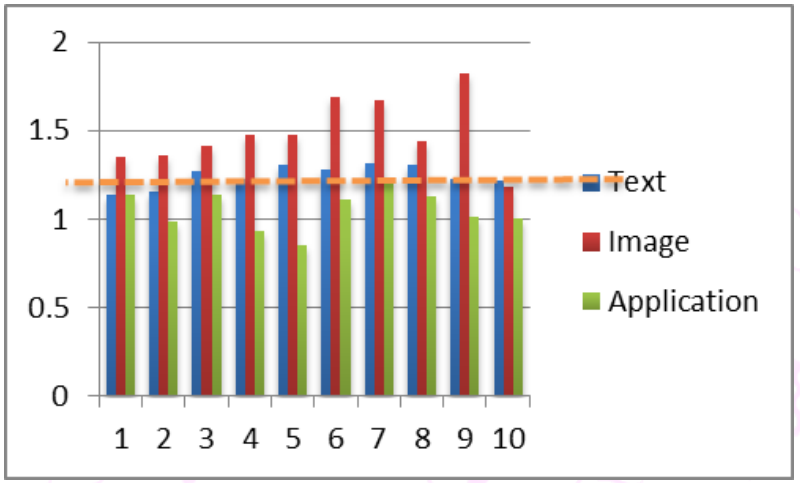


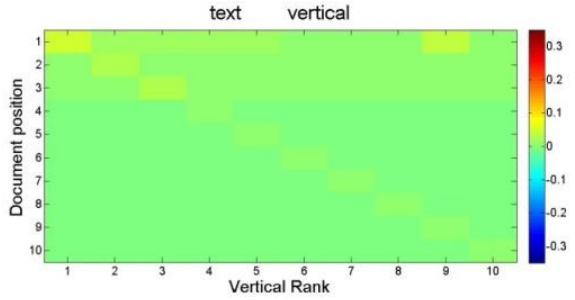
Incorporating presentation bias

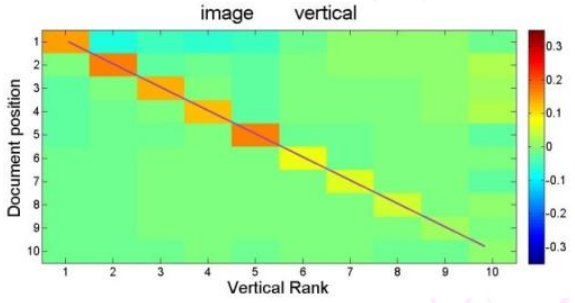
* Presentation bias for vertical results
  * 70% SERPs contain all kinds of vertical results (Sogou, 2012)
  * Certain kinds of vertical results are more attractive than ordinary results (e.g. image/video results)

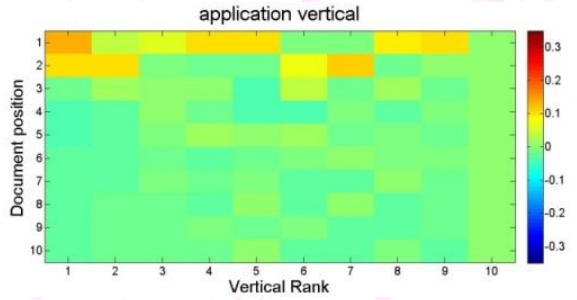
Incorporating presentation bias

- Presentation bias for vertical results
  - Global effect
    - Image results cause global CTR increasing
    - Application results ...
  - Local effect
    - Some results are more attractive









Incorporating presentation bias

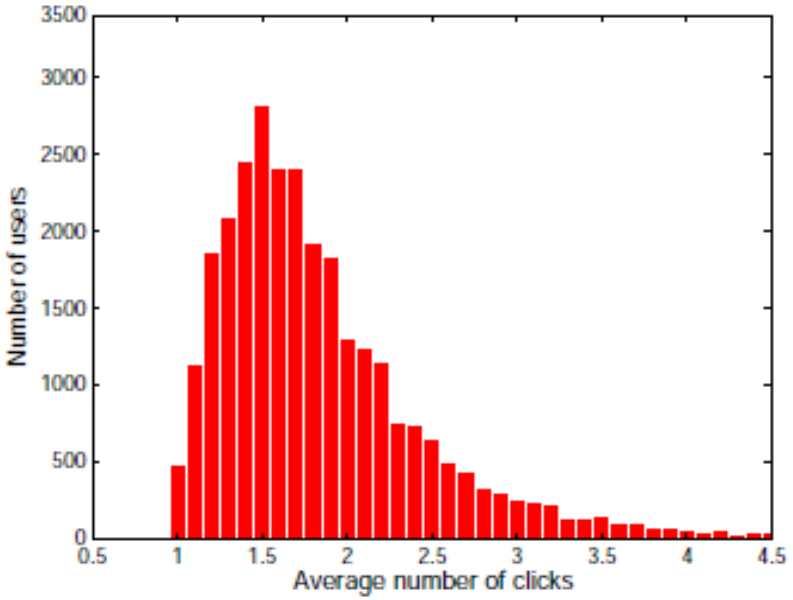
- Presentation bias for vertical results
  - Eye-tracking results show similar findings

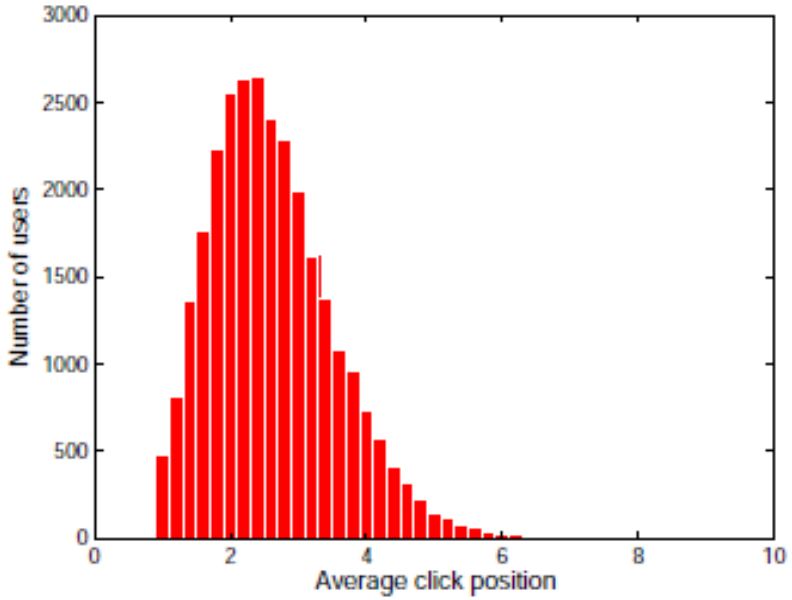


  - How to describe these biases (on-going)
    - Presentation bias model (PBM): attraction bias, global bias, first place bias, sequence bias.

Incorporating user credibility

- User credibility and preference
  - Avg. number of clicks, Avg. position of clicks
  - Search experts, results crawlers, user who has blind faith in search engines, …





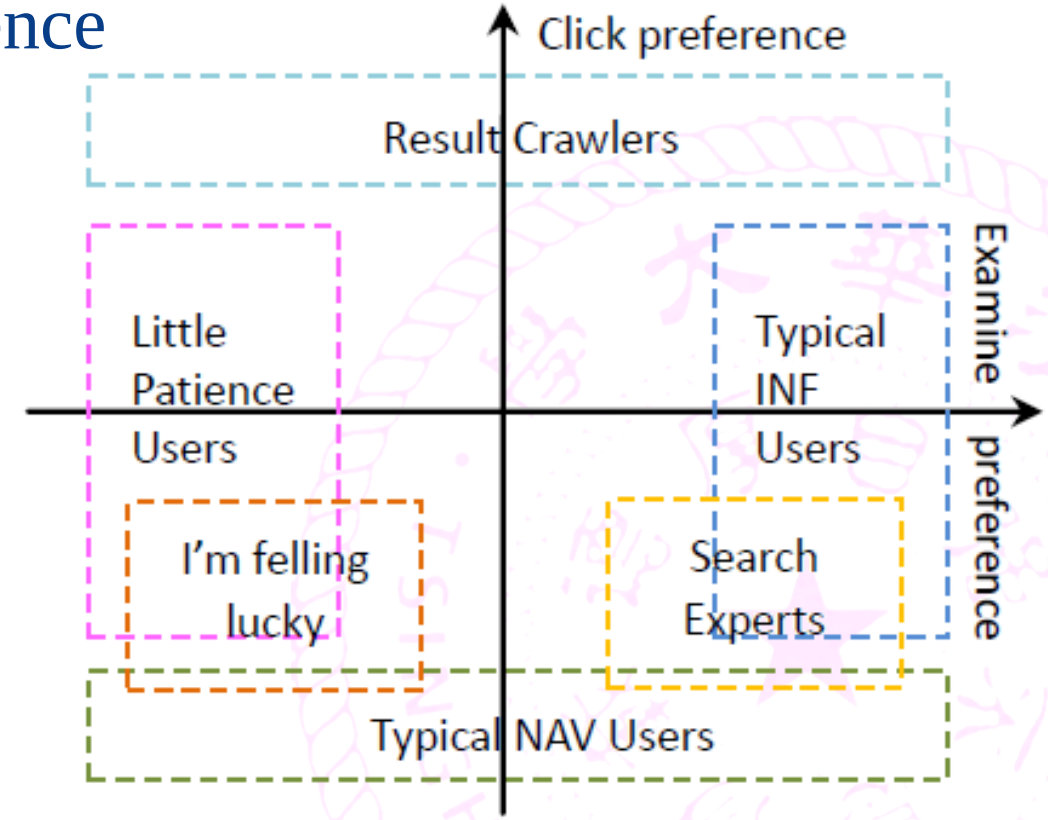
Incorporating user credibility

* How to describe user preference
  * Examination preference

$$P(E_i = 1|u, t) = \frac{1}{1 + exp(-\alpha_{i,t} - \epsilon_u)};$$

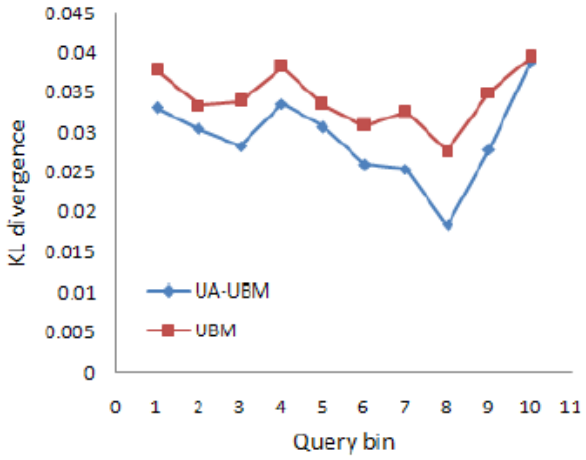
  * Click preference

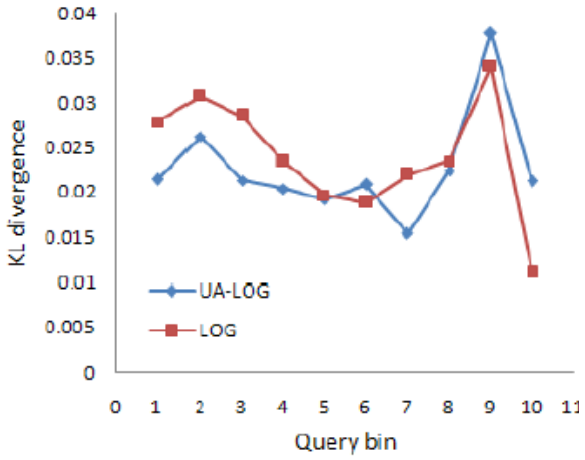
$$P(C_i = 1|E_i = 1, u, q) = \frac{1}{1 + exp(-\beta_{d_i,q} - \gamma_u)};$$

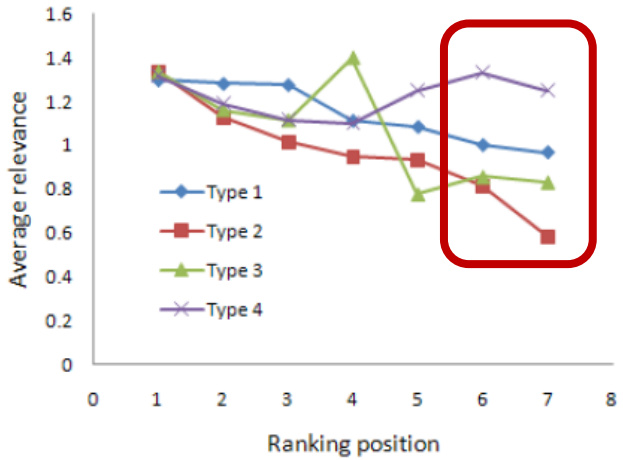


Incorporating user credibility

- Performance Evaluation
  - Prediction of search user behaviors
    - Better than UBM/Cascade/logistic models
  - Prediction of relevance from feedback information
    - Works even better for lower-ranked results







Thank you

Any comments?

Welcome to visit our homepage

http://www.thuir.cn/