

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

Yiqun Liu

Associate Professor, Tsinghua University

Beijing, China



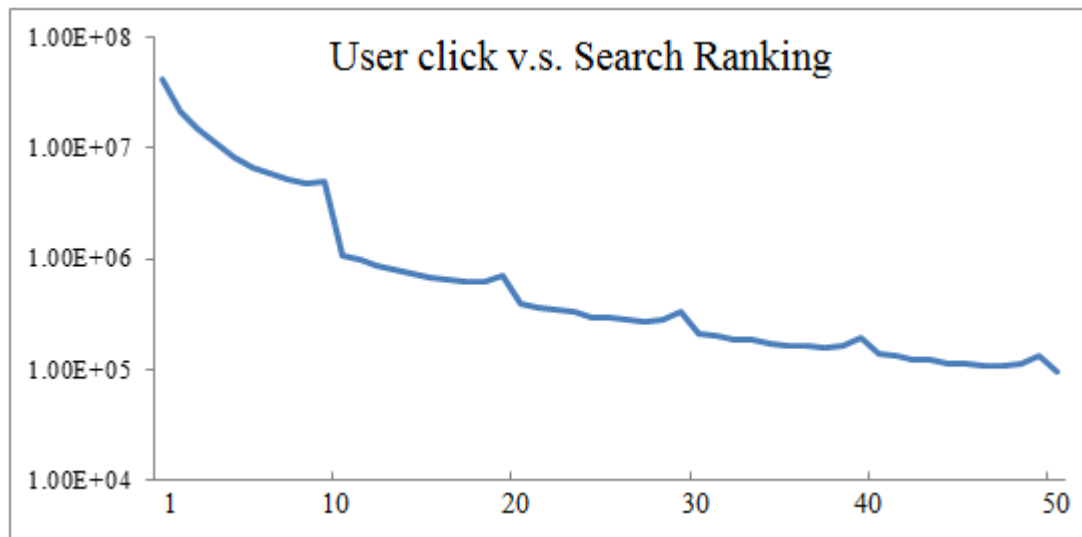
Search Engine Ranking

- * 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 naive 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





Information Retriever @ Tsinghua University

Relevance Feedback

* Possible problem: presentation bias

Search results for '清华大学计算机系' (Tsinghua University Computer Department) on Sogou. The results are heavily biased towards the department's official website and related academic resources, with many links to '清华大学计算机系' and '清华大学计算机系'.

Search results for '郭德纲' (Gou Denggang) on Sogou. The results are heavily biased towards Gou Denggang's official website and related entertainment resources, with many links to '郭德纲' and '郭德纲'.

Search results for 'womendai' (Women's Day) on Sogou. The results are heavily biased towards 'womendai' and related resources, with many links to 'womendai' and 'womendai'.

* Possible problem: user behavior credibility

Search results for 'miserable failure' on Google. The results show a list of links, including 'George W. Bush Biography of the 43rd President of the United States' and 'Jimmy Carter Short biography from the official White House site'. The results are biased towards the White House website and related resources.

Search results for 'liar' on Google. The results show a list of links, including 'Tony Blair - Biography' and 'X X X LIAR - POWER OVER TEMPORARY EUPH'. The results are biased towards Tony Blair's biography and related resources.

Constructing Click Models

- * Examination hypothesis to avoid position bias

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

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

- * 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):

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

- * Other models: DBM, CCM, ...





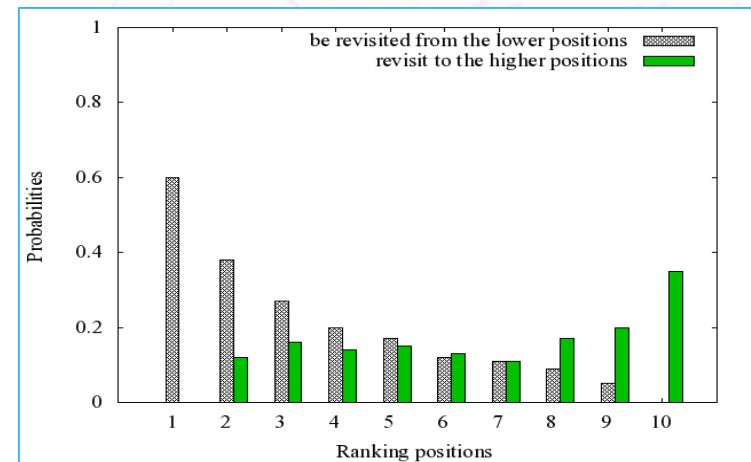
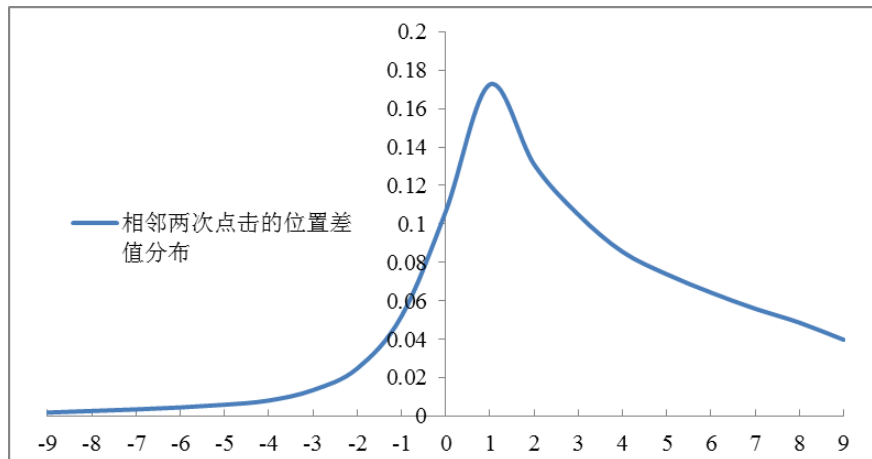
Constructing Click Models

- * 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



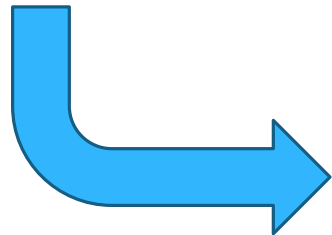
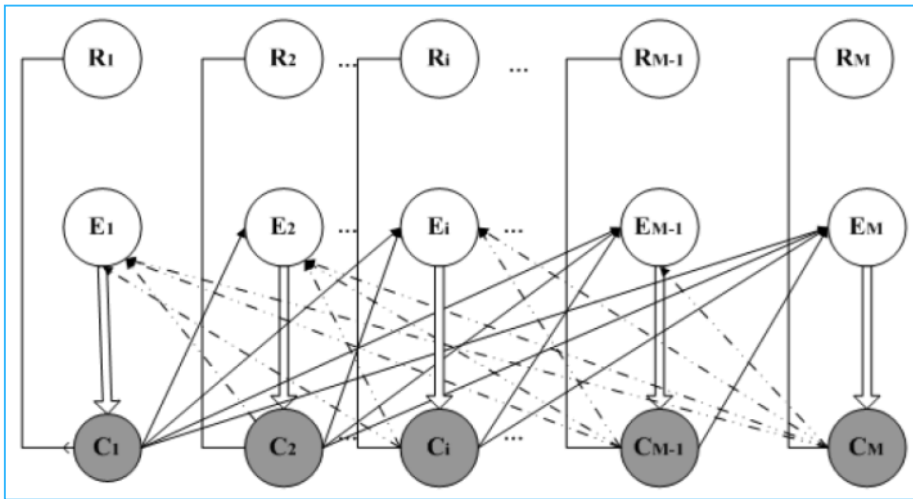
Incorporating Revisiting Behaviors

- * 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



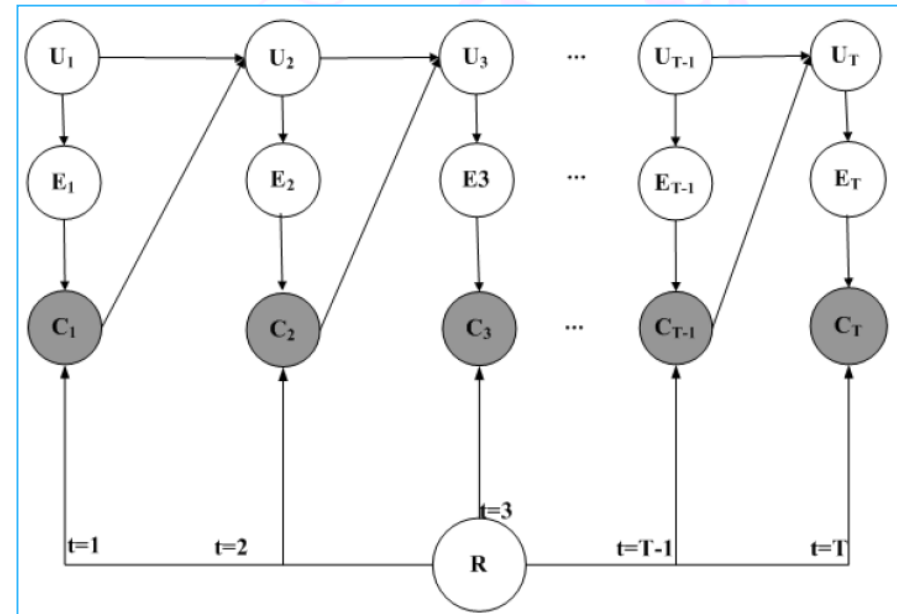
Incorporating Revisiting Behaviors

* THCM: From ranking sequence to time sequence



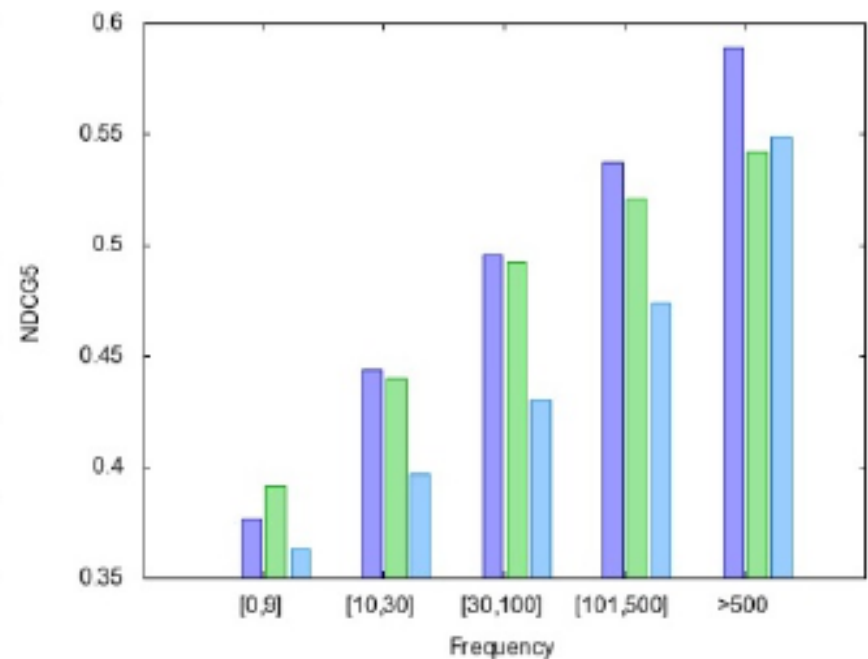
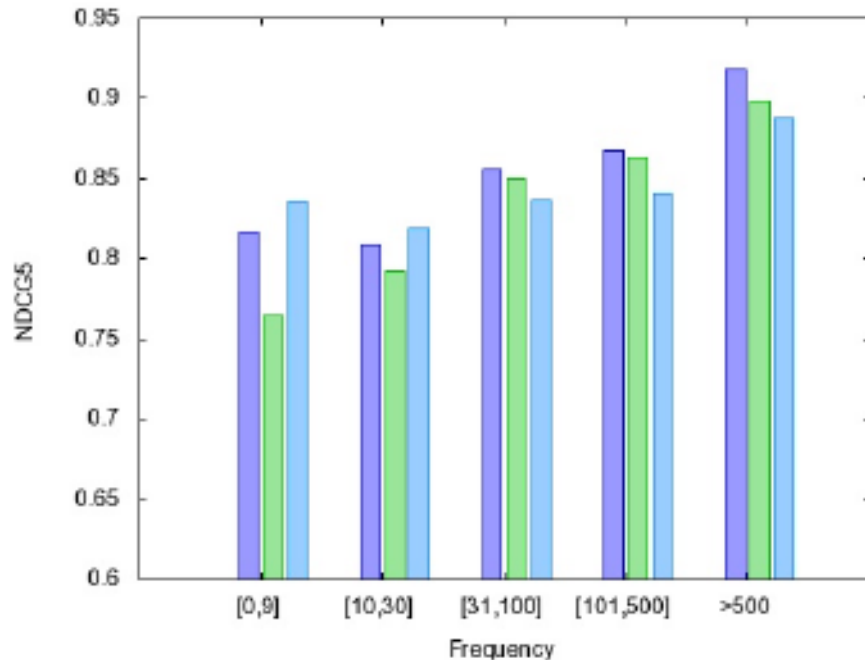
Forward event: $P(E_{t(i+1)} = 1 | E_{ti} = 1) = \alpha$

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



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)

美国大选日期 百度知道

问：美国大选日期

答：美国大选的全国选民投票在选举年11月份的第一个星期一后的第一个星期二举行（2008年是11月4日），这一天被称为总统大选日。所有美国选民都到指定地点进行投票，...

百度知道 - zhidao.baidu.com/question/70101271 - 2008-09-30 - 快照 - 预览

2012美国什么时候选举 百度知道 2012-2-7

2008年美国总统大选日期 百度知道 2008-9-15

2012年美国总统大选的时间是几月 百度知道 2011-8-28

[更多站内相关问题>>](#)

Text

仓木麻衣图片的图片搜索结果-搜狗图片



搜狗图片 - pic.sogou.com/ - 2012-10-5

Image

最新搜狗拼音输入法官方免费下载



搜狗拼音

软件大小：26.85MB

软件版本：6.2.0.7817 更新时间：2012-09-27

运行环境：XP/Vista/Win7/Win8

[官方下载](#)

搜狗拼音输入法 - pinyin.sogou.com - 2012-10-5

Application

搜狗拼音输入法 6.2.0.7817 正式版 下载 - 输入法 - 天空软件...

新增功能：1、Windows 8支持——首家全面支持Win8的输入法 2、图片表情——打字出表情,让你的打字更生动 3、彩虹字——缤纷的字体,全新输入体验 4、火星文——好玩的火星文,输...

天空软件站 - www.skycn.com/soft/27159.html - 2012-9-28 - 快照 - 预览

Traditional

Incorporating presentation bias

* Presentation bias for vertical results

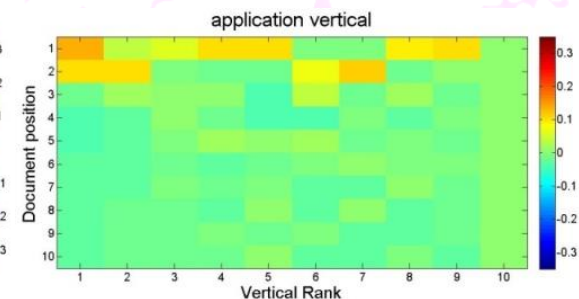
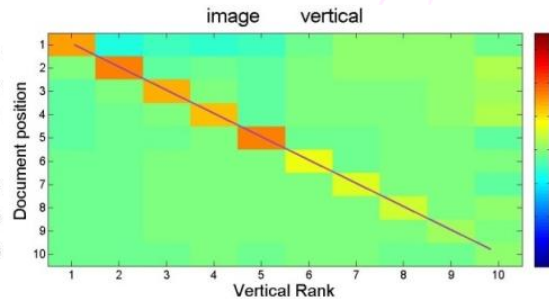
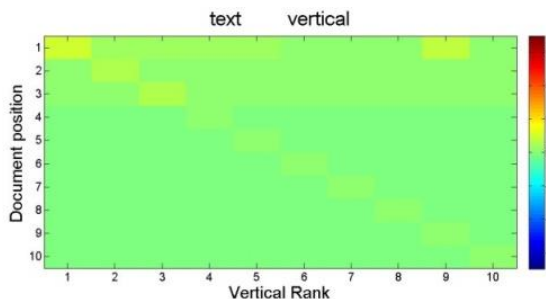
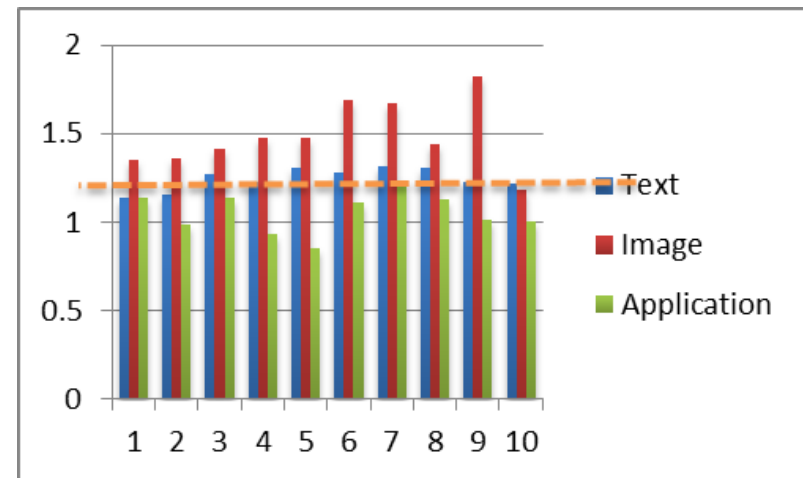
* Global effect

* Image results cause global CTR increasing

* Application results ...

* Local effect

* Some results are more attractive





Information Retriever @ Tsinghua University

Incorporating presentation bias

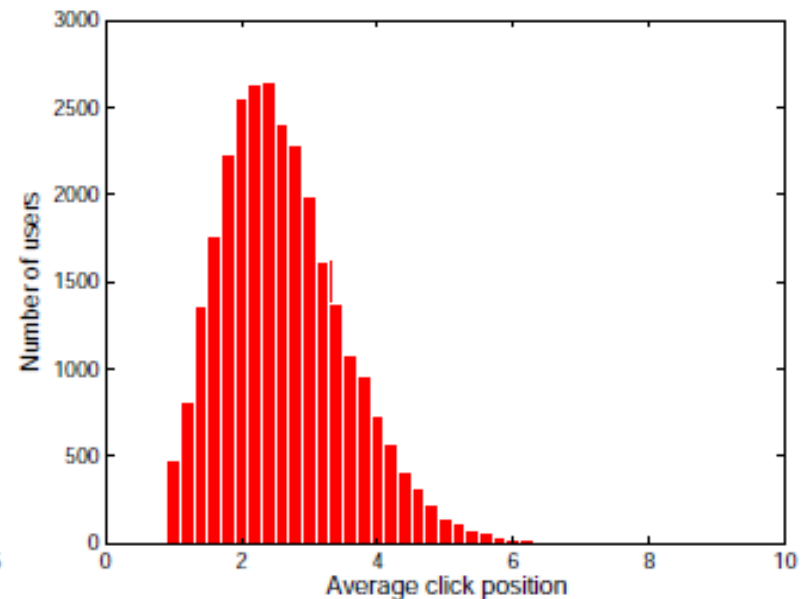
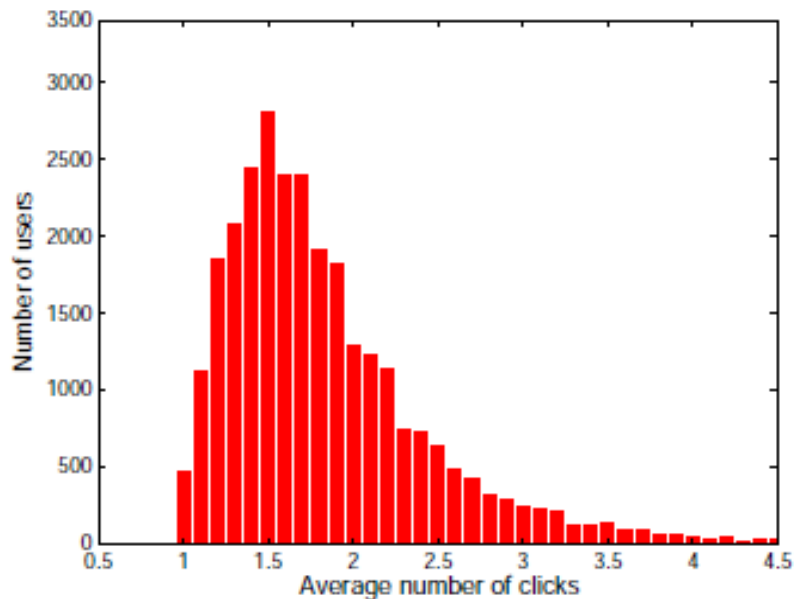
- * Presentation bias for vertical results
- * Eye-tracking results show similar findings



- * 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

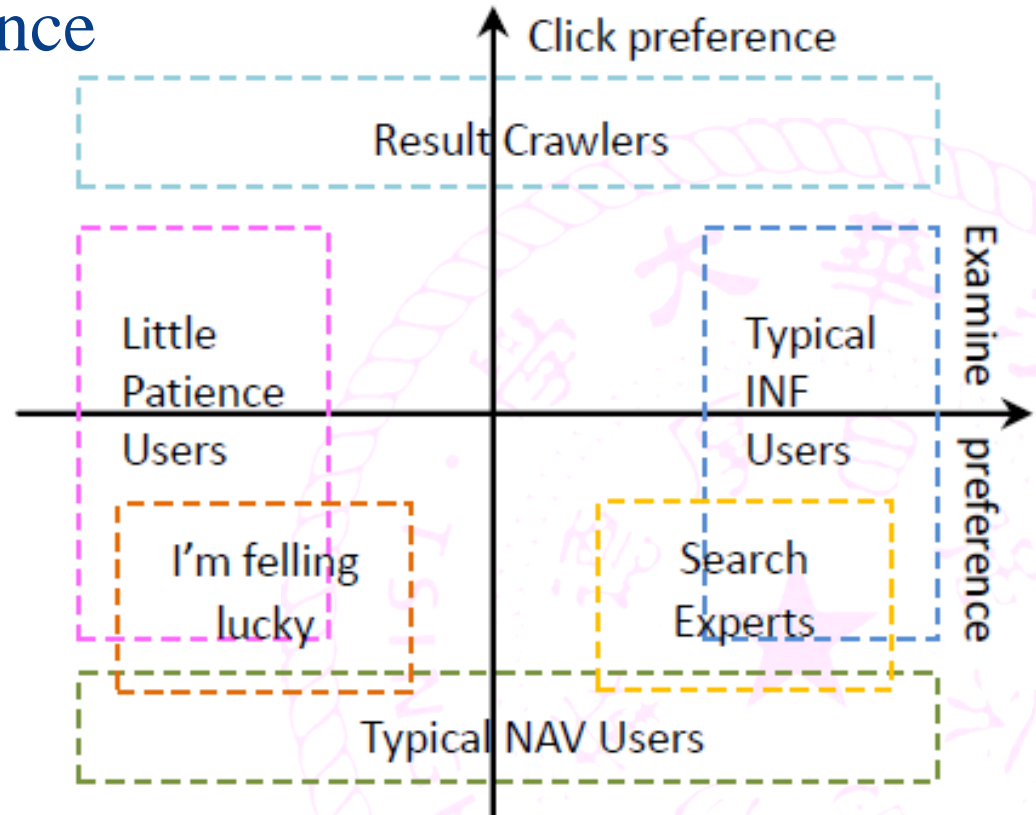
$$P(E_i = 1|u, t)$$

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

* Click preference

$$P(C_i = 1|E_i = 1, u, q)$$

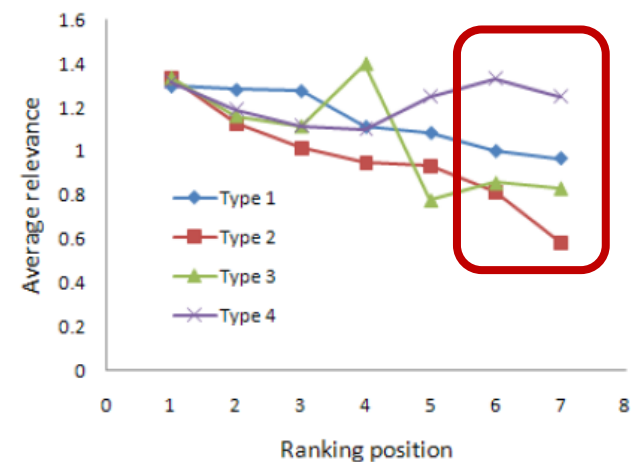
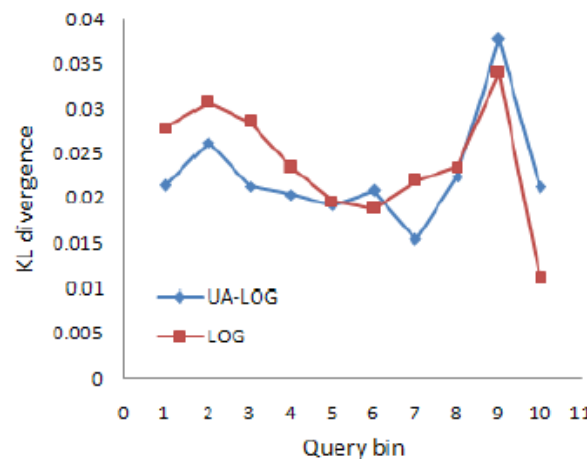
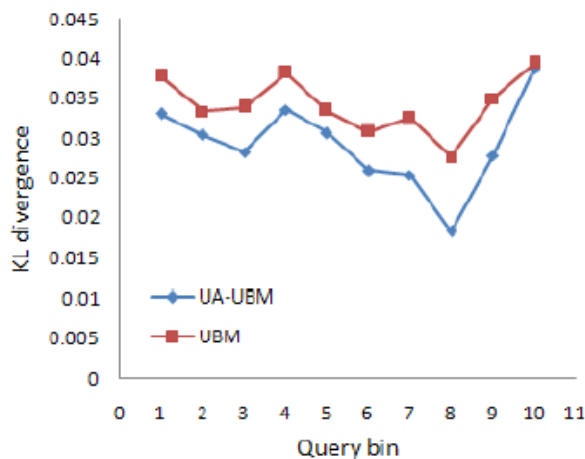
$$= \frac{1}{1 + \exp(-\beta_{d_i,q} - \gamma_u)};$$



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/>

