

# Click Chain Model in Sponsored Search

Presented by WANG Wei

The Chinese University of Hong Kong

# Outline

- Background and motivation
- Click Chain Model
- My recent and future work

# Adapt ranking to user clicks

ALL RESULTS

1-9 of 49 results · [Advanced](#) · [Safe Search Strict](#)

## [KDD 2011: 17th ACM SIGKDD Conference on Knowledge Discovery and ...](#)

Welcome to **KDD-2011**: August 21-24, 2011 Grand Manchester Hyatt, San Diego, CA  
[www.kdd2011.com](http://www.kdd2011.com)

## [KDD2011 \(KDD2011\) on Twitter](#)

**KDD2011 (KDD2011)** is on Twitter. Sign up for Twitter to follow **KDD2011 (KDD2011)** and get their latest updates  
[twitter.com/KDD2011](https://twitter.com/KDD2011)

## [Overview | VLDB 2011](#)

VLDB is a premier annual international forum for data management and database researchers, vendors, practitioners, application developers, and users. The conference will feature research talks, tutorials, demonstrations, and workshops. It will cover current issues in data management, database and information systems research.  
[vldb.org/2011](http://vldb.org/2011)

## [KDD-2011 Call for Papers](#)

Industry Practice Expo: [www.kdd.org/kdd2011/indexpo.shtml](http://www.kdd.org/kdd2011/indexpo.shtml) Key Dates: Aug 21-24, 2011 **KDD-2011** Conference June 30, 2011 KDD Cup Competition Ends May 13, 2011 Paper Acceptance Mar 15, 2011 KDD Cup Competition Begins Mar 1, 2011 KDD Cup Registration Opens ([www.kdd.org/kdd2011/kddcup.shtml](http://www.kdd.org/kdd2011/kddcup.shtml))  
[www.kdnuggets.com/2011/01/kdd-2011-cfp.html](http://www.kdnuggets.com/2011/01/kdd-2011-cfp.html)

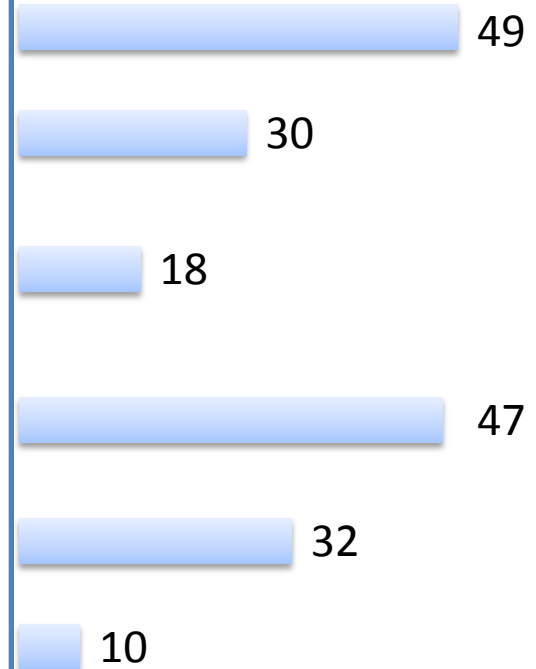
## [KDD 2011: KDD Cup](#)

KDD-Cup 2011: Recommending Music Items based on the Yahoo! Music Dataset  
[www.kdd.org/kdd2011/kddcup.shtml](http://www.kdd.org/kdd2011/kddcup.shtml)

## [Calit2 : California Institute for Telecommunications and ...](#)

Guest Speaker: Pdraic Smyth, UC Irvine and Calit2 Professor, Computer Science and **KDD2011** Program Co-Chair, et al. Date: August 21st, 2011 to August 24th, 2011 Time: All day Location: Grand Manchester Hyatt Hotel, UC San Diego Host: Padhraic Smyth, UCI [event website]  
[www.calit2.net/events/popup.php?id=1845](http://www.calit2.net/events/popup.php?id=1845)

## # of clicked



**Motivation – Click Data Are Valuable**

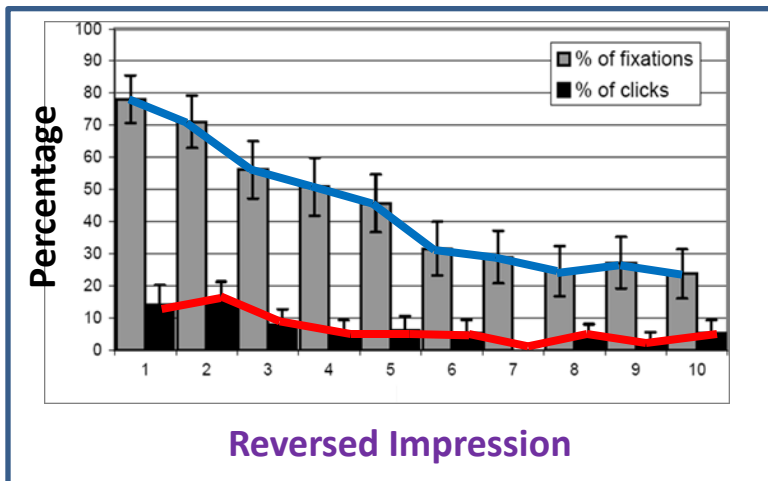
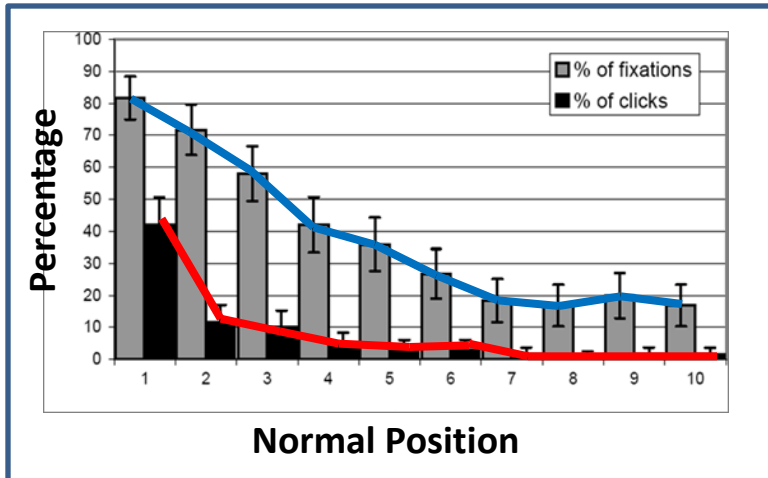
# Web Search Click Log

- Auto-generated data keeping important information about search activity.

Query	kdd2011	
Position	URL	Click
1	<a href="http://www.kdd2011.com/">http://www.kdd2011.com/</a>	1
2	<a href="http://twitter.com/#!/KDD2011">http://twitter.com/#!/KDD2011</a>	0
3	<a href="http://vldb.org/2011/">http://vldb.org/2011/</a>	0
4	<a href="http://www.kdnuggets.com/2011/01/kdd-2011-cfp.html">http://www.kdnuggets.com/2011/01/kdd-2011-cfp.html</a>	1
5	<a href="http://www.kdd.org/kdd2011/kddcup.shtml">http://www.kdd.org/kdd2011/kddcup.shtml</a>	0
6	<a href="http://www.calit2.net/events/popup.php?id=1845">http://www.calit2.net/events/popup.php?id=1845</a>	0

One session

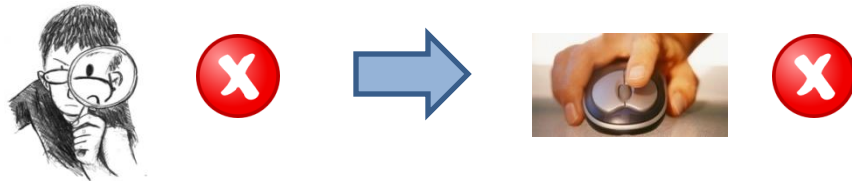
# Click Position Bias Problem



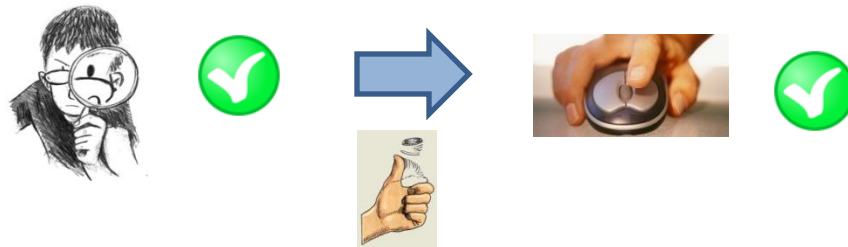
- Given: a set of web search click logs
- **Estimate relevance:** measures how good a URL is with regard to the information need of the query/user

# Examination Hypothesis

- A document must be examined before a click.

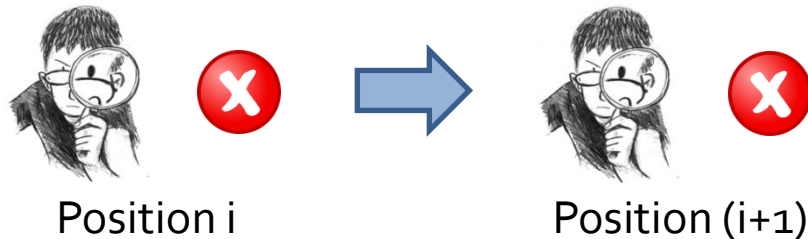


- The (conditional) probability of click upon examination depends on document relevance.



# Cascade Hypothesis

- The first document is always examined.
- First-order Markov property:
  - Examination at position  $(i+1)$  depends on examination and click at position  $i$  only
- Examination follows a strict linear order:



# Algorithms

- Given the click data  $S^{1:U}$ , compute the posterior distribution  $p(R_i | S^{1:U})$ .

$$p(R_i | S^{1:U}) \approx c \times p(R_i) \prod_{u=1}^U p(S_i | R_i)$$

The equation is annotated with three callout boxes:

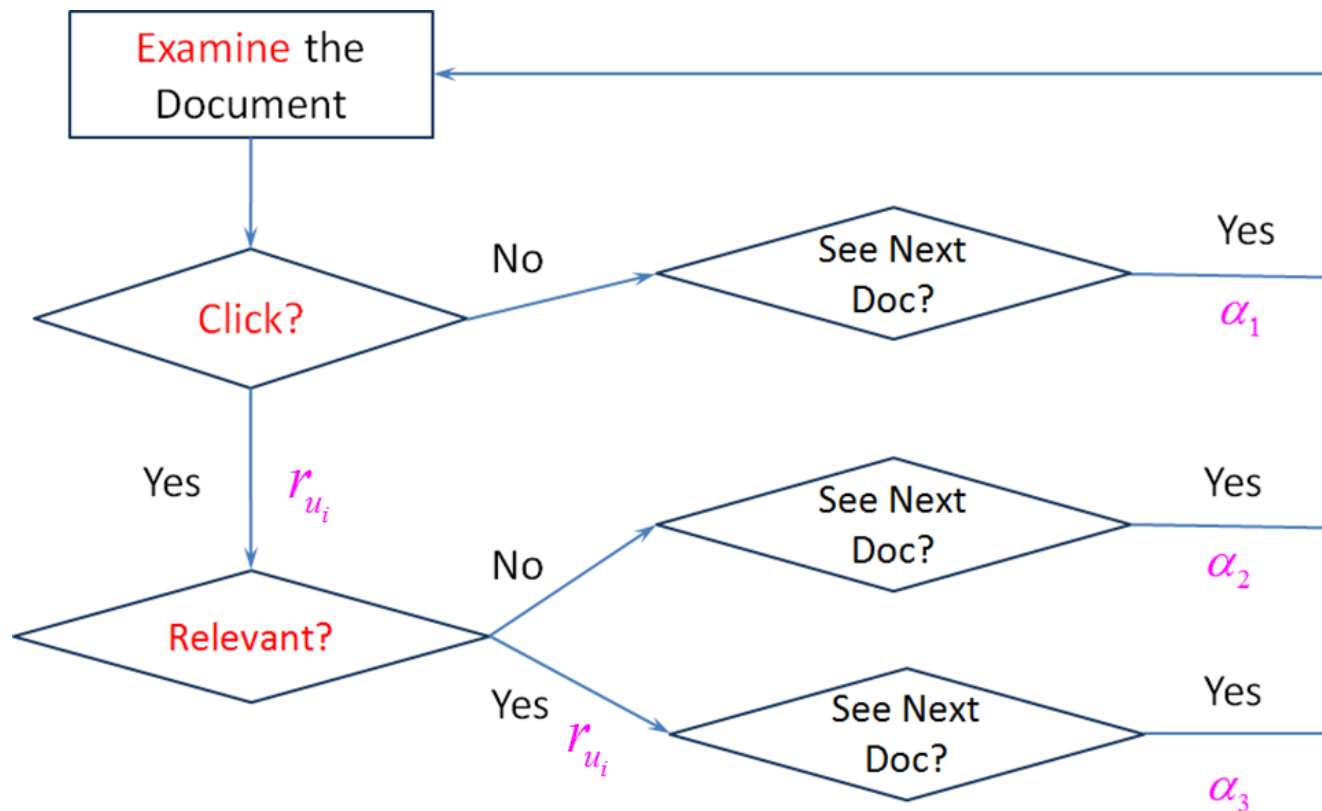
- A box labeled "constant" points to the term  $c$ .
- A box labeled "prior probability" points to the term  $p(R_i)$ .
- A box labeled "conditional probability for each session" points to the product term  $\prod_{u=1}^U p(S_i | R_i)$ .

- Clicks in query sessions as conditionally independent random variables
- Predict  $p(S_i | R_i)$  using Click Chain Model



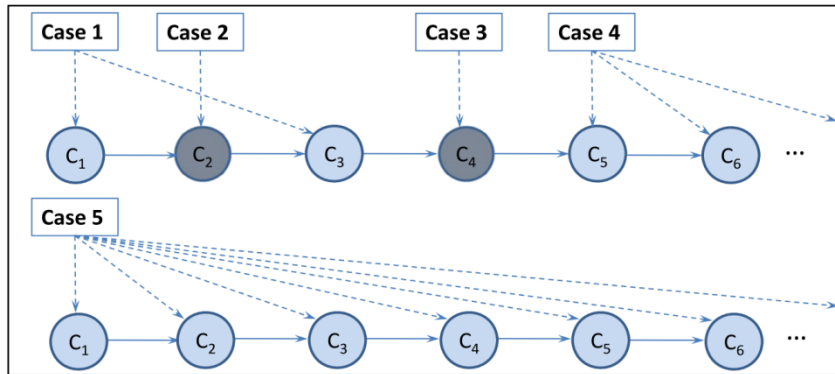
# Click Chain Model

- The full picture



# Estimating $p(S_i | R_i)$

- Focus on last clicked position:  $l = \arg \max_j \{C_j = 1\}$



Case	Conditions	Results
1	$i < l, C_i = 0$	$1 - R_i$
2	$i < l, C_i = 1$	$R_i(1 - (1 - \alpha_3/\alpha_2)R_i)$
3	$i = l$	$R_i \left(1 + \frac{\alpha_2 - \alpha_3}{2 - \alpha_1 - \alpha_2} R_i\right)$
4	$i > l$	$1 - \frac{2}{1 + \frac{6 - 3\alpha_1 - \alpha_2 - 2\alpha_3}{(1 - \alpha_1)(\alpha_2 + 2\alpha_3)} (2/\alpha_1)^{(i-l)-1}} R_i$
5	No Click	$1 - \frac{2}{1 + (2/\alpha_1)^{i-1}} R_i$

# Estimating Relevance

- Recall:

$$p(R_i | S^{1:U}) \approx c \times p(R_i) \prod_{u=1}^U p(S_u | R_i)$$

- $p(R_i)$  : uniform distribution
- **Bayes estimation** for relevance:

$$\int_0^1 \frac{p(r)dr}{c} = 1 \implies c = \int_0^1 p(r)dr$$

$$E(r) = \int_0^1 r \frac{p(r)}{c} dr = \frac{\int_0^1 rp(r)dr}{\int_0^1 p(r)dr}$$

# Parameter Estimation

- Maximum likelihood estimation:

$$\begin{aligned}\ell(\alpha) = & N_1 \log \alpha_1 + N_2 \log \alpha_4 + N_3 \log(6 - 3\alpha_1 - \alpha_4) \\ & + N_5 \log(1 - \alpha_1) - (N_3 + N_5) \log(2 - \alpha_1) \\ & - N_1 \log 2 - (N_2 + N_3) \log 6\end{aligned}$$

$$\alpha_1 = \frac{3N_1 + N_2 + N_5 - \sqrt{(3N_1 + N_2 + N_5)^2 - 8N_1(N_1 + N_2)}}{2(N_1 + N_2)}$$

and

$$\alpha_4 = \frac{3N_2(2 - \alpha_1)}{N_2 + N_3}$$

where  $\alpha_4 \triangleq \alpha_2 + 2\alpha_3$ , and N means the number for 5 cases.

# My Recent Work

query

網頁 知識+ 圖片 新聞 討論區 Video 分類廣告 更多

how to study english

搜尋

YAHOO!  
雅虎香港

顯示:  香港網頁優先  中文網頁  全球網頁

how to study english 搜尋結果約554,000,000個, 以下為1 - 10個 - 0.55 秒

贊助網站

贊助網站

## 英國文化協會 英語課程

幼兒至中小及成人英語課程, 國際認可專業英語導師親身任教, 面對面全方位學習。

[www.britishcouncil.org.hk](http://www.britishcouncil.org.hk)

## Prime English

免費英語示範課程助你掃走港式英語, 全港參加人數至今累積超過25,000人。

[www.primeenglish.com.hk](http://www.primeenglish.com.hk)

## 十五個星期英語達標

按照設計進度, 就能達到職業英語標準! 持續進修基金資助八成學費。

[www.ICanLearn.hk/workplaceenglish](http://www.ICanLearn.hk/workplaceenglish)

## Hello Education

學了多年英語還未能純熟運用? 我們提供實用及針對性英文課程, 助您真正掌握所學。

[www.helloeducation.hk](http://www.helloeducation.hk)

## Study English - How to Study English Effectively

This short introduction to [how to study English](#) effectively provides helpful hints and exercises to help you [study English](#) well. ... going to [study English](#) on topics that focuses on ...

[esl.about.com/od/intermediateenglish/a/study\\_english.htm](http://esl.about.com/od/intermediateenglish/a/study_english.htm) - 網頁紀錄 - 更多此站結果

## How to Study English 7 Tips and Ideas | UK Student ...

Use our top 7 tips for [how to study English](#). ... 3 Responses to "How to Study English 7 Tips and Ideas" ... is [English](#) native speaker, and chat with them really help my [English](#) skill, ...

[adviceforyou.org.uk/blog/.../how-to-study-english-7-tips-and-ideas](http://adviceforyou.org.uk/blog/.../how-to-study-english-7-tips-and-ideas) - 69k - 網頁紀錄 - 更多此站結果

Sponsored search results

## 學好英語, 國際英文雜誌入手

多間資優學校推薦, 學好英語, 從閱讀國際權威英文雜誌開始, 現在優惠價訂閱。

[www.newspaper.com.hk](http://www.newspaper.com.hk)

## 活學研習中心

位於銅鑼灣, 各科補習班現正招生, 英語課程透過角色扮演增加學習興趣。

[www.learningstepstudio.com](http://www.learningstepstudio.com)

## 港專英語課程

提供適合一至高中程度的書面及應對技巧英語課程, 歡迎查詢。

[www.hkct-hkcc.edu.hk](http://www.hkct-hkcc.edu.hk)

## 1小時踢走港式英語

唔好再講Open the Light, 三個月學成地道英語, 強化職場競爭力。

[www.modernenglishgroup.com](http://www.modernenglishgroup.com)

## 公文式英文課程

課程循序漸進, 培養孩子主動學習, 適合所有幼兒及中小學生就讀。

[www.kumon.com.cn/hk](http://www.kumon.com.cn/hk)

滄州阿德萊德 優委學府集申世

# Considering the Relevance

- Consider the relevance of a query and its Ads
  - If the Ads are not relevant to the query, it will obtain less click.
- Using Click Chain Model to estimate the relevance

# Experiment

- Dataset:
  - Crawled from Sogou from mainland China
  - Each click log includes 6 fields: query, ad title, ad body, bid terms, bid price and whether click
  - Contains 18010850 sessions
  - $\text{CTR}(\text{Pos.1})=0.02515$
  - $\text{CTR}(\text{Pos.2})=0.009369$
  - $\text{CTR}(\text{Pos.3})=0.006844$

# Dataset Analysis

Query Freq.	#Unique Query	#Session	Avg CTR
1	1695146	1695146	0.0212
2	1058697	1342854	0.0153
3-4	772810	1275067	0.0141
5-8	262555	925065	0.0151
9-17	128304	880444	0.0162
18-32	47981	685582	0.0179
33-221480	48582	4427998	0.0201

Ads Freq.	#Unique Ad	#Impression	Avg CTR
1	26267	26267	0.0257
2	14689	29378	0.0222
3-4	17539	60146	0.0216
5-8	18058	113186	0.0208
9-17	18092	223102	0.0191
18-32	12786	306201	0.0190
33-221480	36365	17252570	0.0178



# Effectiveness and Result

- Effectiveness: log-likelihood
- Base line: predicted by frequency
- Result

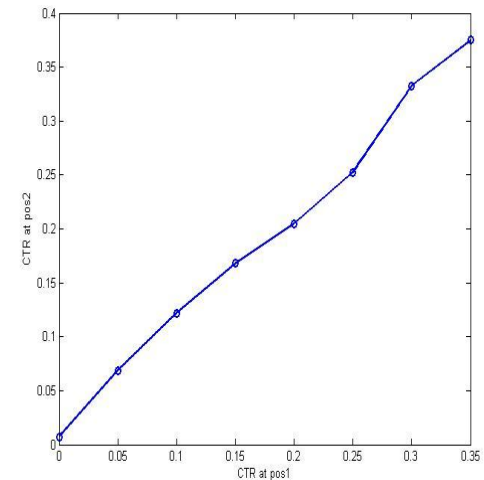
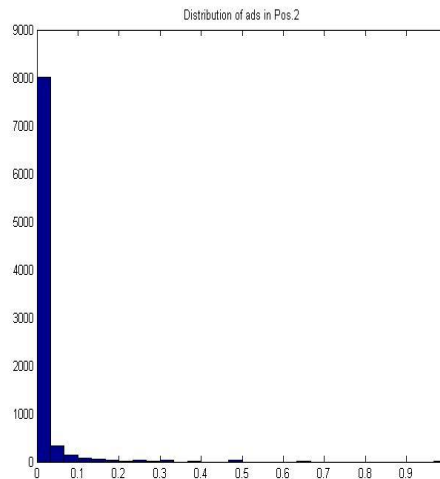
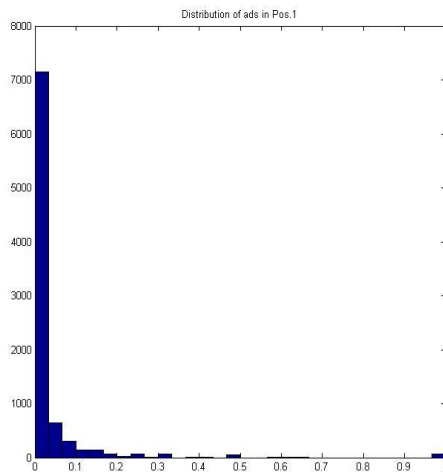
Freq.	1-500	500-1000	1000-5000	5000-10000	10000-15000	15000-20000	>20000
Baseline	-0.2626	-0.218	-0.2331	-0.2261	-0.2021	-0.1992	-0.2049
CCM	-0.2505	-0.2195	-0.2272	-0.219	-0.1946	-0.1941	-0.1964

# My Future Work

- Whether position bias in sponsored is as clear as the general search?
- Whether sponsored search obey the linear sequence?
- Any other models?

# My Future Work

- Some cooperation between ad in pos1 and pos2?



Thanks.