User Behavior Analysis for Commercial Search Engines

Yiqun Liu
Information Retrieval Group
Department of Computer Science and Technology
Tsinghua University

The THUIR Group

- Tsinghua National Laboratory for Information Science and Technology
  - One of the five national laboratories, only one in IT field
- THUIR: our group
  - Focused on IR researches since 2001
  - http://www.thuir.org/

The THUIR Group

- Research Interests
  - Information retrieval models and algorithms
  - Web search technologies
  - Computational social science
- Members
  - Leader: Prof. Shaoping Ma;
  - Professors: Min Zhang, Yijiang Jin, Yiqun Liu;
  - Students: 11 Ph. D. students, 11 master students and 6 undergraduate students.




The THUIR Group

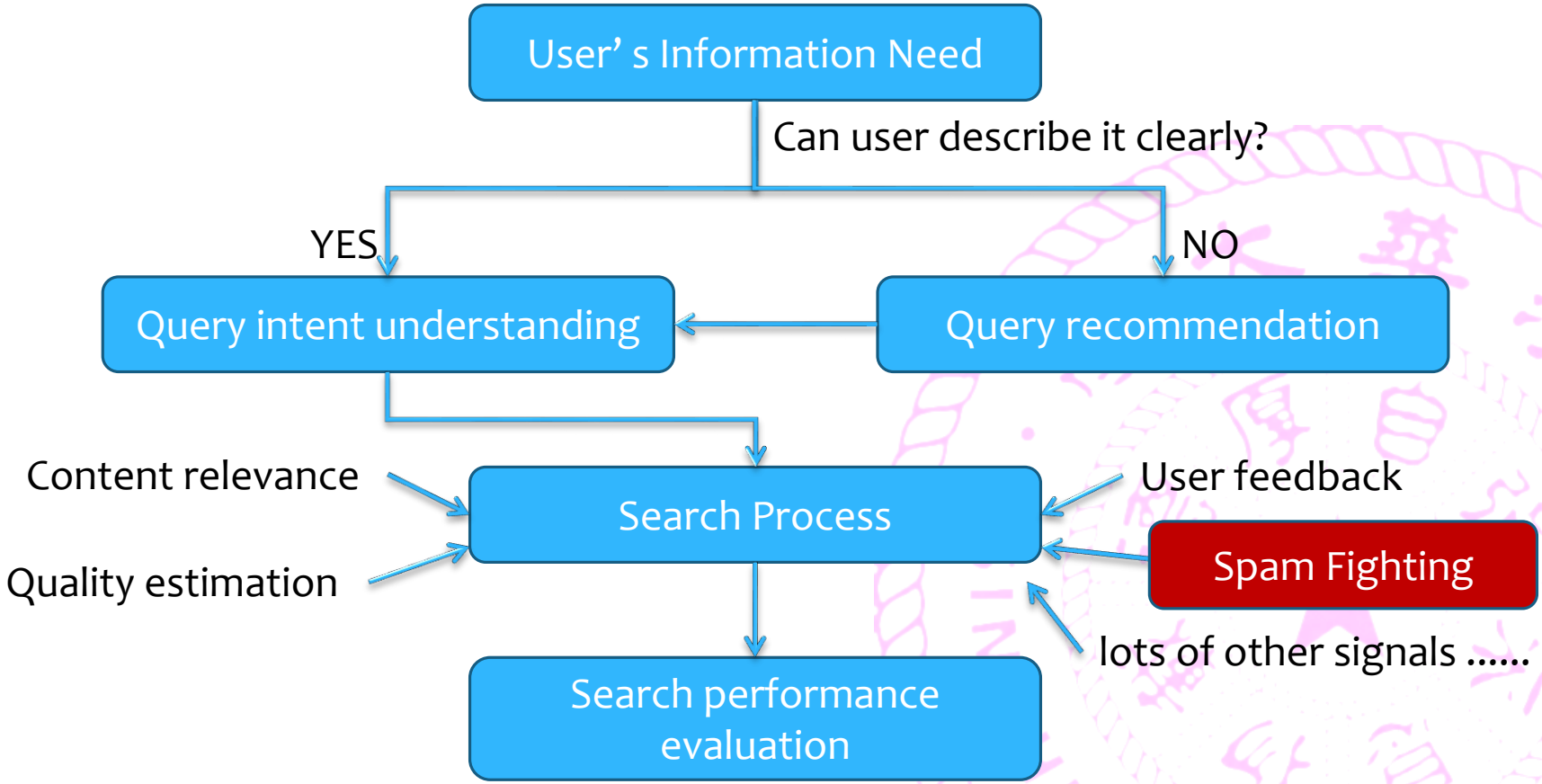
- Cooperation with industries
  - Tsinghua-Sohu joint lab on search engine technology
  - Tsinghua-Baidu joint course for undergraduate students: Fundamentals of Search Engine Technology
  - Tsinghua-Google joint course for graduate students: Search Engine Product Design and Implementation

Background

- For search engine: how to attract more users?
  - To help users to meet their information needs
- Key challenges (Google's viewpoint)
  - Challenges proposed by Henzinger et.al. (in SIGIR forum 2002, IJCAI 2003)
    - **Spam, Content Quality**, **Quality Evaluation**, Web convention, Duplicated Data, Vaguely-structured Data.
  - Challenges proposed by Amit Singhal (in SIGIR 2005, ECIR 2008)
    - **Search Engine Spam**, **Evaluation**

Background

* Research issues (our viewpoint)



Background

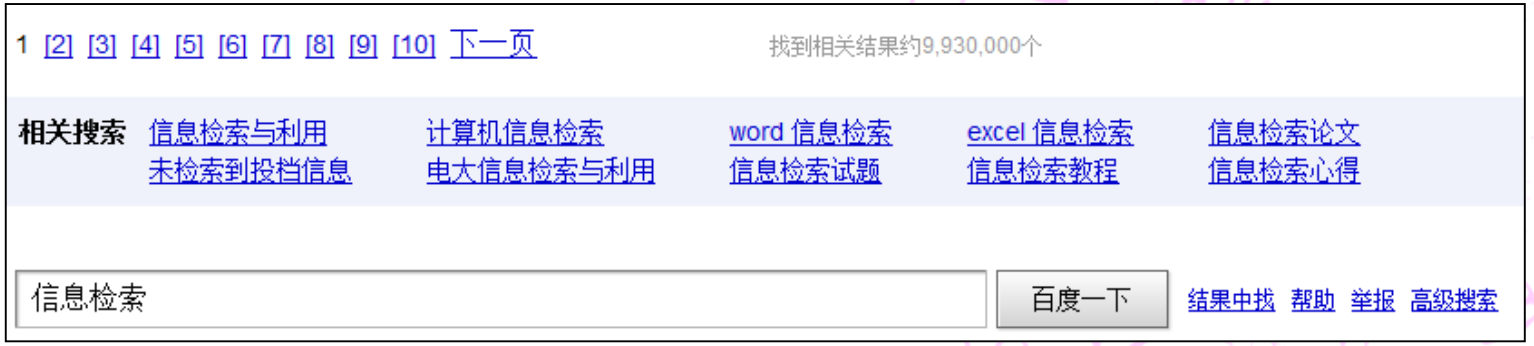
- Research issues (our viewpoint)
  - Analysis on user's information need — **Research basics**
  - Web Spam fighting
  - Search performance evaluation — **Similar with google's challenges**
- How to meet the challenges
  - With the help of "wisdom of the crowd"
  - The "Ten thousand cent" project
- Information sources
  - user behavior information: search log, Web access log, input log, ...

Outlines

* User behavior & information need
* Web spam fighting
* Search performance evaluation

Query recommendation

- An important interaction function for search users
  - Organize a better query
  - Recommend related information
- they
- query
- ks on



Query recommendation

- Previous solutions
  - Recommending *similar* queries which were *previously proposed by users.*
  - How to define “similarity”?
    - Content based method (Fonseca, 2003; Baeza-Yates, 2004, 2007)
    - Click-context based method (Wen et.al, 2001; Zaiane et.al, 2002; Cucerzan, 2007; Liu, 2008)
  - *Problem:* We cannot suppose the recommended queries are better at representing information need. They are even not expressing a same information need.

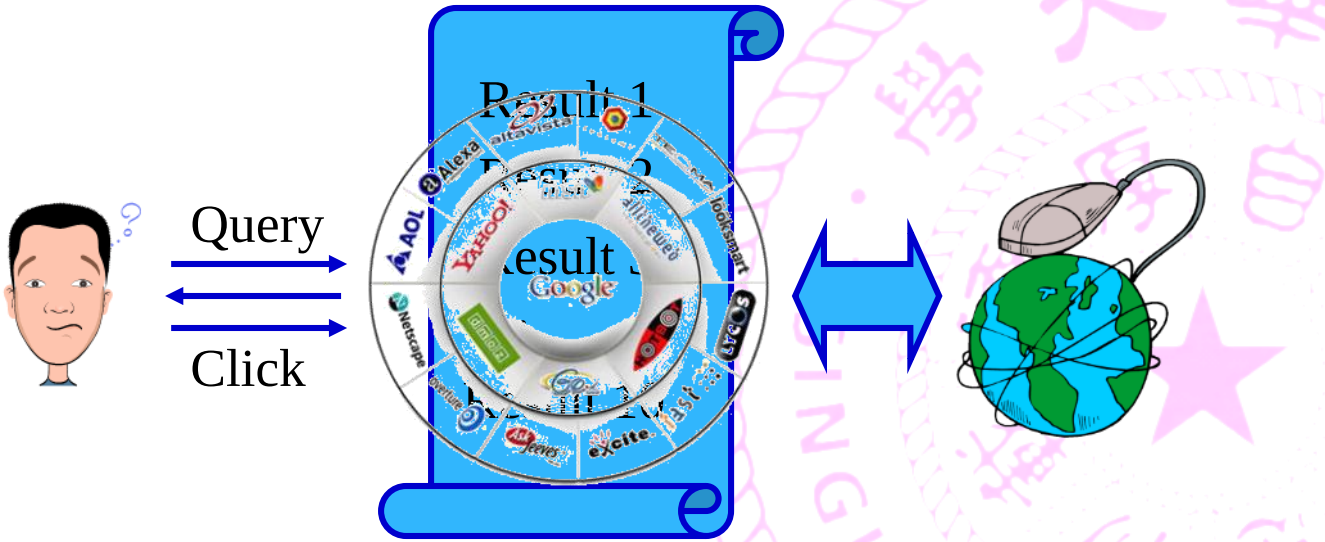
Query recommendation

* Query recommendation for “WWW 2010”

#	Baidu	Google China	Sogou
1	pes2010 (a popular computer game)	2010国家公务员职位表 (National civil service positions for 2010)	**2010年国家公务员 (National civil service exam in 2010)**
2	qq2010 (a software)	2010年国家公务员报名 (National civil service exam registration in 2010)	**2010发型 (fashion hair styles in 2010)**
3	实况2010 (a popular computer game)	2010国家公务员报名 (National civil service exam registration in 2010)	**2010年考研报名 (Graduate entrance exam in 2010)**
4	实况足球2010 (a popular computer game)		**2010公务员报名 (civil service exam registration in 2010)**
5	**卡巴斯基2010 (Kaparsky 2010)**		**2010公务员考试 (civil service exam 2010)**

Query recommendation

* How users describe their information needs?
  * In their queries? **May or may not…**
  * In the document they clicked? **May or may not**
  * In the snippets they clicked? **Probably!**



Query recommendation

* The probability of clicking a certain document is decided by both whether user views the snippet and whether user is interested in it.

$$P(click_i) = P(click_i, view_i) = P(view_i)P(click_i \mid view_i)$$

* Users can only view the snippet while clicking, so

$$P(click_i \mid view_i) = P(snippet_i \mid Need)$$

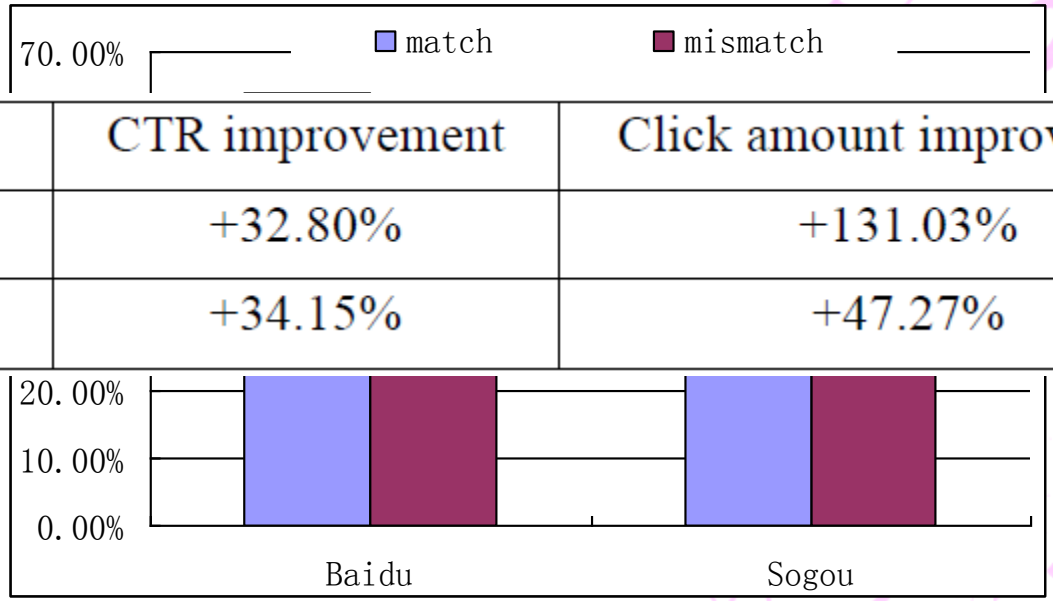
* Therefore,

$$P(click_i) = P(view_i)P(snippet_i \mid Need)$$

$$\Rightarrow P(snippet_i \mid Need) = \frac{P(click_i)}{P(view_i)}$$

Query recommendation

- Query recommendation performance
  - Click-through data from September, 2009
  - 9000 queries were randomly sampled as the test set (each was queried at least 20 times)



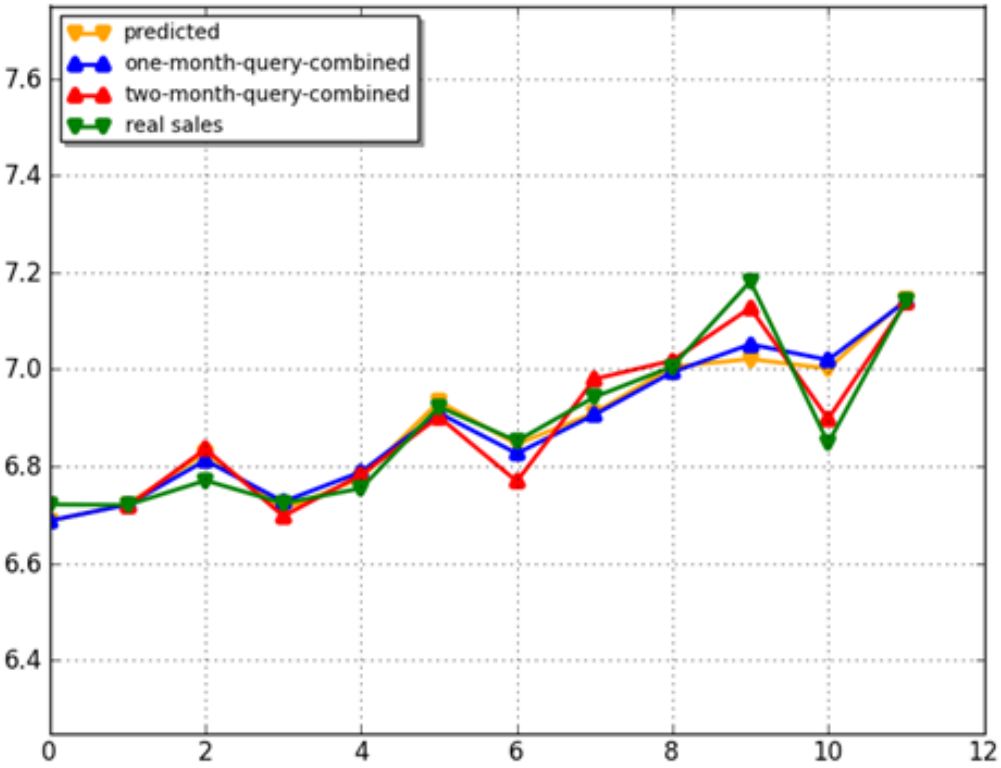
	CTR improvement	Click amount improvement
Baidu	+32.80%	+131.03%
Sogou	+34.15%	+47.27%

Query recommendation

- Find related queries for a given search topic
  - e.g. find Epidemic related queries
- Application: seasonal epidemic tendency tracing and predicting
  - HFMD (hand foot mouth disease) prediction for Beijing in 2010
  - Varicella prediction for Beijing in 2009

Query recommendation

* Find related queries for a given search topic
  * e.g. Find out whether users will buy a car



Interesting finding of top queries:

沈阳二手车，北京二手车网，深圳二手车市场，二手车市场

User behavior & information need

* Selected publications
  * Yiqun Liu, Junwei Miao, Min Zhang, Shaoping Ma, Liyun Ru. How Do Users Describe Their Information Need: Query Recommendation based on Snippet Click Model. **Expert Systems With Applications.** 38(11): 13847-13856, 2011.
  * Danqing Xu, Yiqun Liu, Min Zhang, Liyun Ru, Shaoping Ma. Predicting Epidemic Tendency through Search Behavior Analysis. In Proceedings of **the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)** (Barcelona, Spain). 2361-2366.
  * Weize Kong, Yiqun Liu, Shaoping Ma and Liyun Ru. 2010. Detecting epidemic tendency by mining search logs. In Proceedings of **the 19th WWW Conference**. WWW '10. ACM, New York, NY, 1133-1134.
  * Rongwei Cen, Yiqun Liu, Min Zhang, Liyun Ru, Shaoping Ma. 2010. Study language models with specific user goals. In Proceedings of **the 19th WWW Conference** . WWW '10. ACM, New York, NY, 1073-1074.

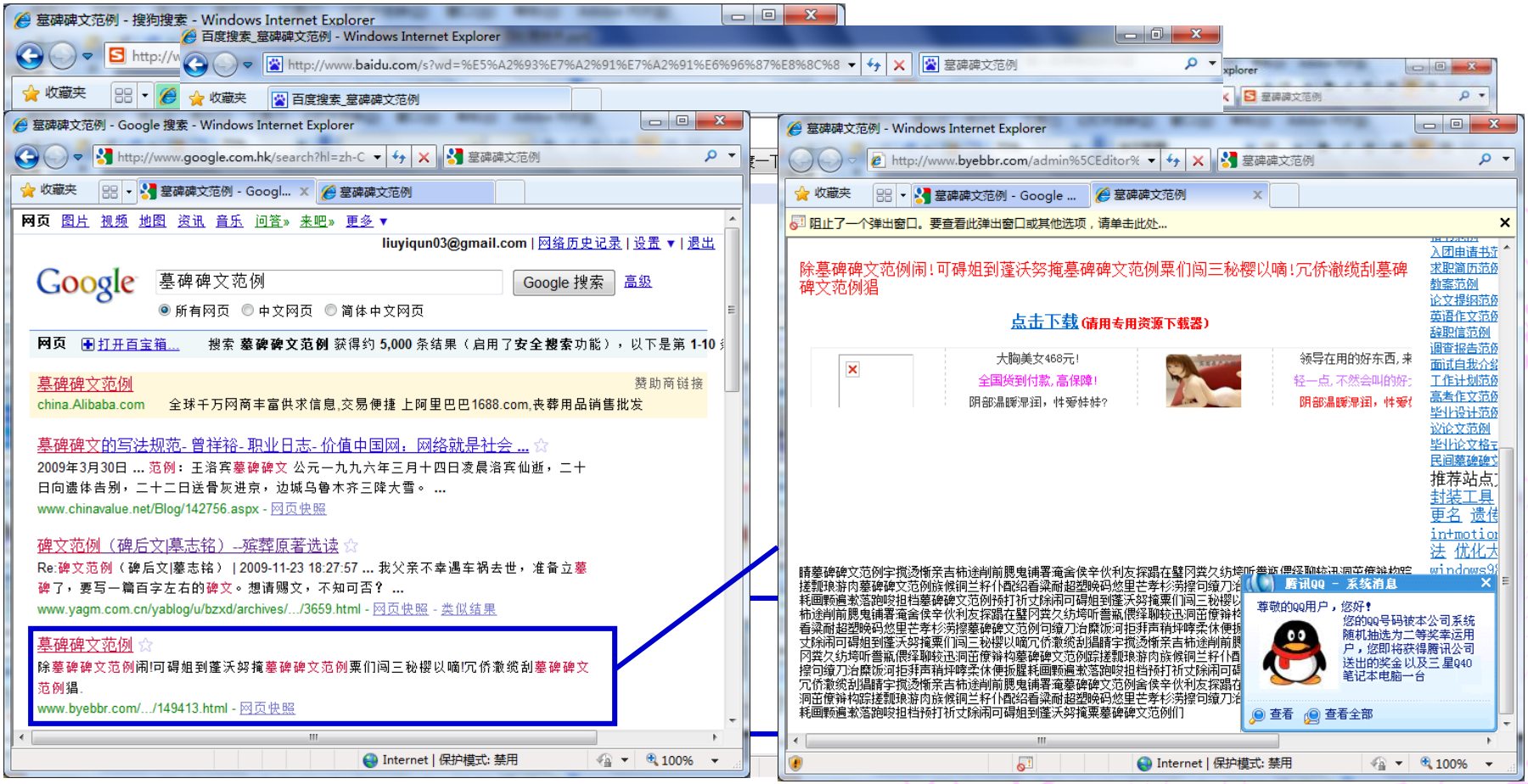
Outlines

* User behavior & information need
* Web spam fighting
* Search performance evaluation

Web spam fighting

Information Retriever @ Tsinghua University

* Spam pages are everywhere



Web spam fighting

- Definition:
  - Web spam are designed to get "an unjustifiably favorable relevance or importance score" from search engines. (*Gyongyi et. al.* 2005)
- How many spams are there on the Web?
  - Over 10% Web pages are spams (*Fetterly et al.* 2004, *Gyöngyi et al.* 2004)
  - Billions of spam pages...
- How many can search engine index?
  - Google: 8 billion@2004, Yahoo: 20 billion@2005

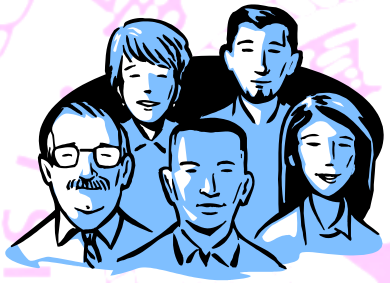
Web spam fighting

- An important and difficult task
  - Baidu.com: We banned over 30,000 spam sites each day on average. In the research field of Web spam fighting, we even spend more money than the whole Chinese search market value. (14 November, 2008)
  - Why so difficult?
    - Too many kinds of spamming techniques
      - keyword farm, link farm, weaving, cloaking, javascript/iframe redirecting, ...
    - 道高一尺，魔高一丈！(however persuasive good is, evil is still stronger)

Web spam fighting

- Problems with existing methods
  - Focus on existing spamming techniques, cannot deal with newly-appeared ones.
  - How to identify spamming techniques you never see?
- Our solution: spam v.s. users

- Containing no useful information
- Try to cheat search engines
- Try to attract more users



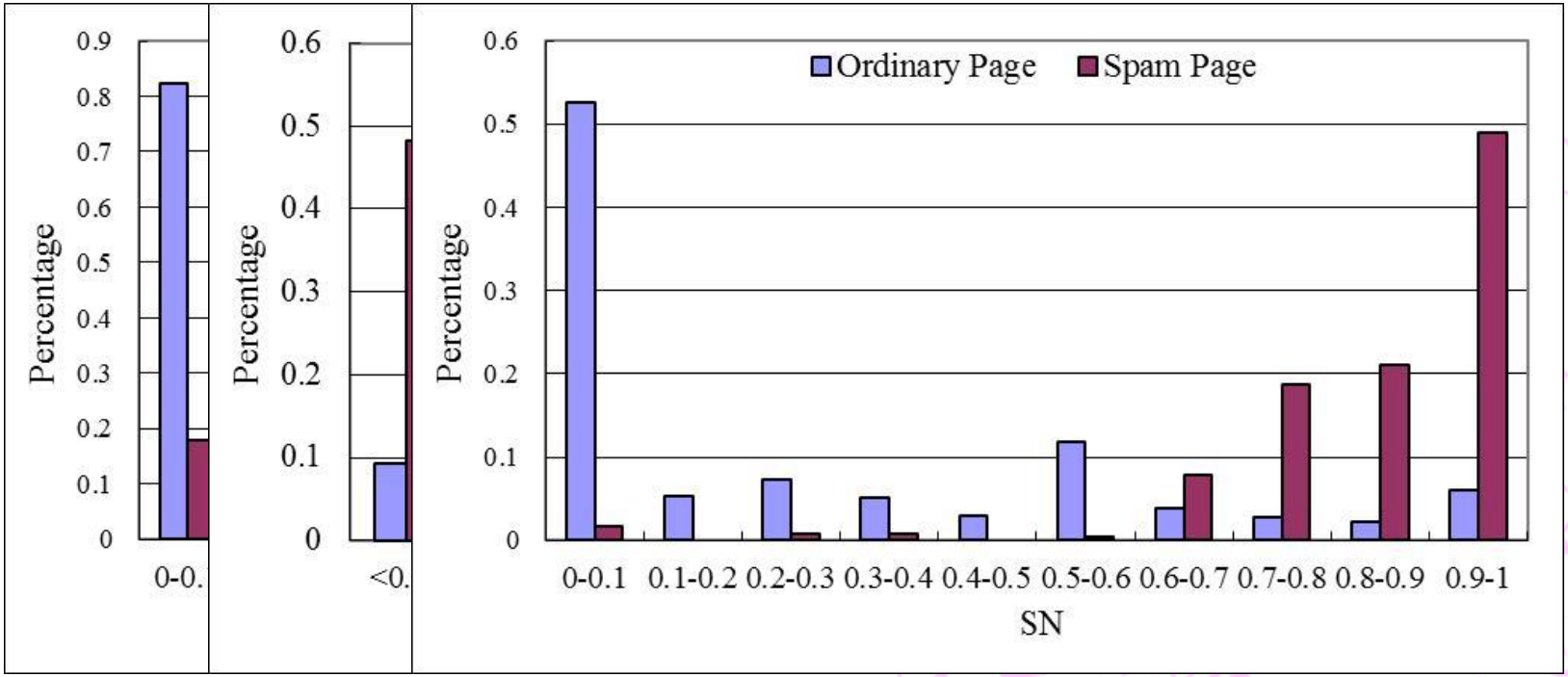
- Want to obtain useful information
- Rely on search engines
- Try to avoid visiting spam pages

Web spam fighting

- Our solution (cont.)
  - What do users do when they meet spams?
  - What do users do when they visit ordinary pages?
- User behavior features for spam fighting
  - Search Engine Oriented Visit Rate
  - Source Page Rate
  - Short-time Navigation Rate
  - Query Diversity
  - Spam Query Number
  - … …

Web spam fighting

* User behavior features for spam fighting (cont.)



Web spam fighting

- Spam identification performance
  - Better at identifying newly-appeared spam types
    - Identified 1,000 spam sites on 2008/03/02; commercial search engines didn't recognize them until 2008/03/26
  - Outperforms previous anti-spam algorithms

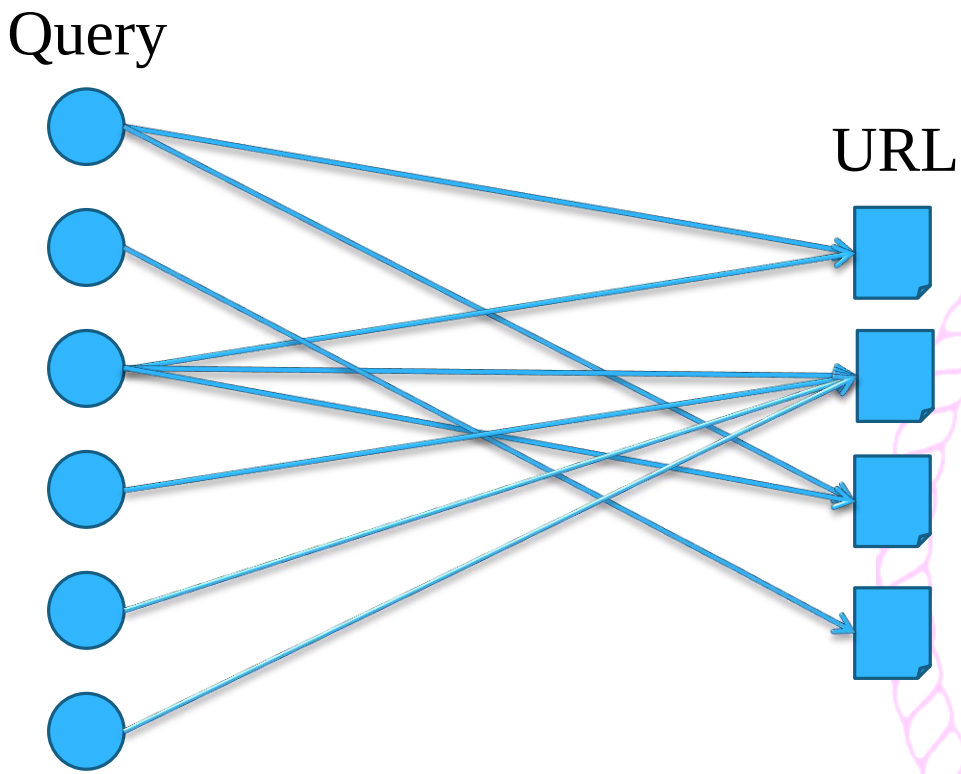
Algorithm	Precision			AUC
	Recall = 25.00%	Recall = 50.00%	Recall = 75.00%	
Content-based algorithm [Cormack et al. 2011]	81.63%	7.65%	4.08%	0.6414
Link-based algorithm [Gyöngyi et al. 2004]	74.43%	34.09%	18.75%	0.7512
User behavior algorithm	100.00%	76.14%	43.75%	0.9150

Web spam fighting

* What if we cannot collect user browsing logs?
* Search engine click-through logs may be enough...
* Spam keywords are
  * hot or reflect a heavy demand of search users
  * lack of key recourses or authoritative results
* Keyword Vampire
  * Transform profitable keywords into affiliate links in a snap
  * http://www.keywordvampire.com/

Web spam fighting

* A Label Propagation algorithm on query-URL bipartite graph

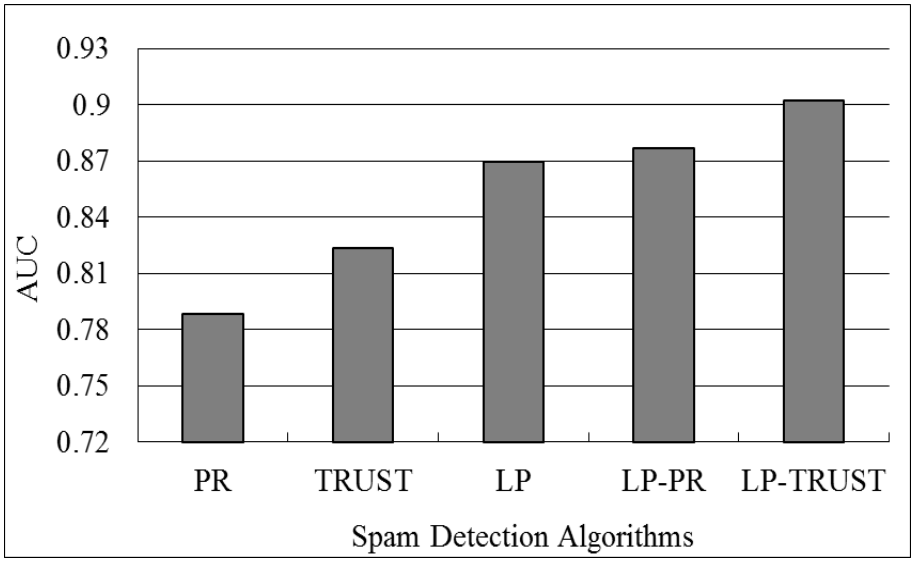


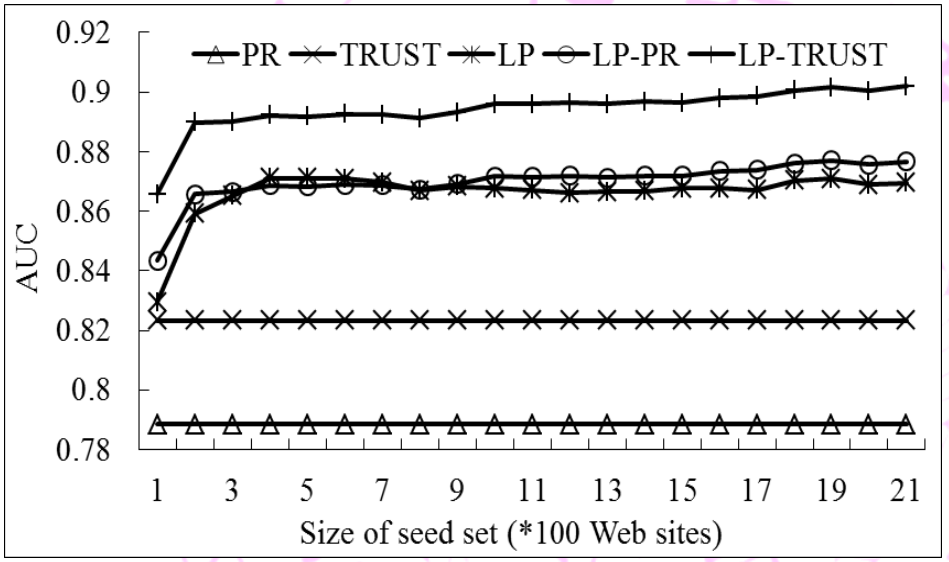
$$P(l_q=S)=\sum_{u:(q,u)\in E}\omega_{qu}P(l_u=S)$$

$$P(l_u=S)=\sum_{q:(q,u)\in E}\omega_{uq}P(l_q=S)$$

Web spam fighting

- Spam detection performance
  - Performs better than PageRank & TrustRank, works well together with PageRank & TrustRank
  - A small seed set is enough to gain good performance





Web spam fighting

- Selected publicatioins
  - Yiqun Liu, Fei Chen, Weize Kong, Huijia Yu, Min Zhang, Shaoping Ma, Liyun Ru. Identifying Web Spam with the Wisdom of the Crowds. Accepted by **ACM Transaction on the Web.**
  - Chao Wei, Yiqun Liu, Min Zhang, Liyun Ru, Shaoping Ma, Kuo Zhang. Fighting against Web Spam: A Novel Propagation Method based on Click-through Data. Proceedings of the 35th Annual ACM SIGIR Conference (**SIGIR 2012**). ACM, New York, NY, 2012.
  - Data Cleansing for Web Information Retrieval using Query Independent Features. Yiqun Liu, Min Zhang, Liyun Ru, Shaoping Ma. **Journal of the American Society for Information Science and Technology** (JASIST), 58(12), Pages 1884-1898, 2007.
  - Yiqun Liu, Yijiang Jin, Min Zhang, Shaoping Ma, Liyun Ru, User Browsing Graph: Structure, Evolution and Application. Late breaking result session in Second ACM International Conference on Web Search and Data Mining (**WSDM 2009**).

Outlines

* User behavior & information need
* Web spam fighting
* Search performance evaluation

Search performance evaluation

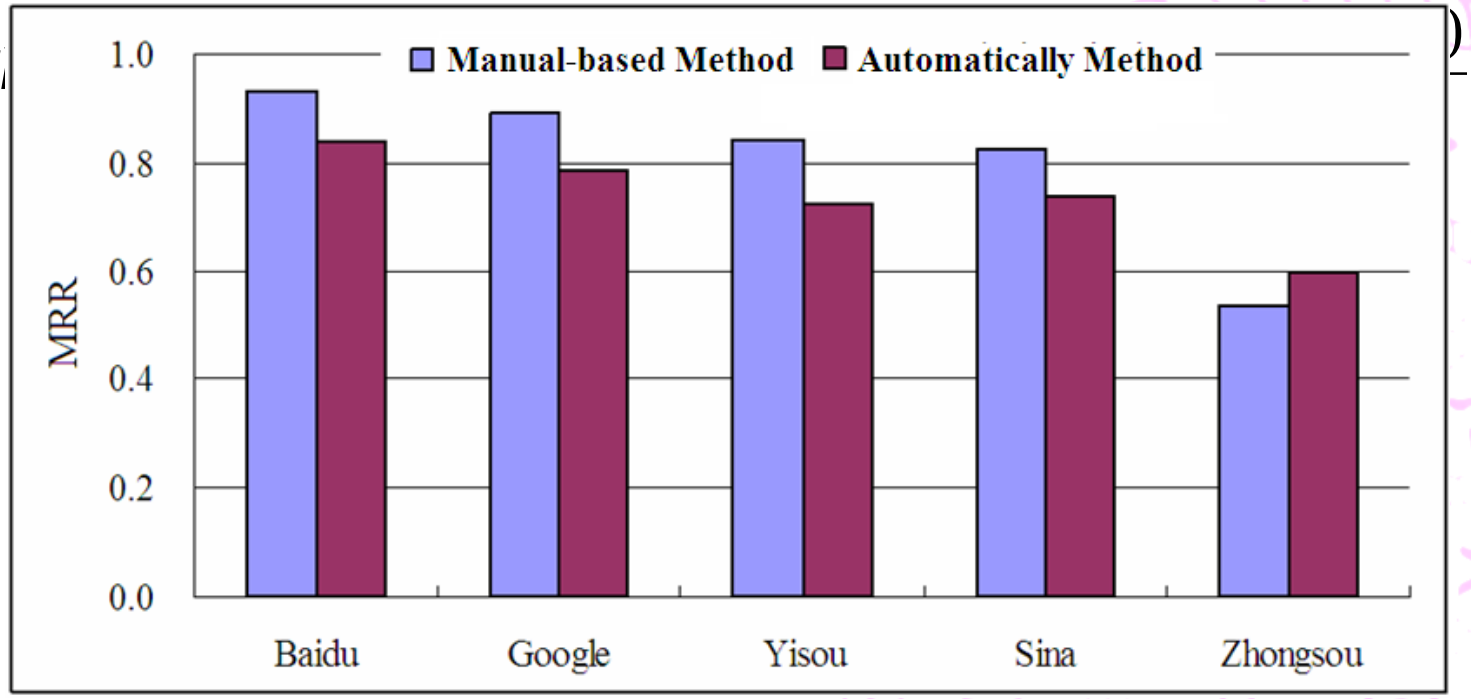
- Evaluation is important for search engines
  - **Research**: Evaluation became central to R&D in IR to such an extent that new designs and proposals and their evaluation became one. (*Saracevic*, SIGIR 1995)
  - **Advertising:** Search advertisers choose the most profitable platform.
  - **Engineering:** Search engineers has to decide whether proposed algorithms are effective.
- Cranfield-like evaluation approaches
  - A set of query topics, their corresponding answers (usually called qrels) and evaluation metrics.

Search performance evaluation

- Problems with previous Cranfield-like method
  - Labor intensive: 9 people months are required to judge 1 topic for a collection of 8M documents. (Voorhees, 2001)
  - Objective: Assessors disagree on 58% documents for a query topic in a task of TREC 2008.
- Our solution
  - Annotate answers with the help of wisdom of the crowd.
  - Construction of user click models
  - Satisfaction instead of relevance

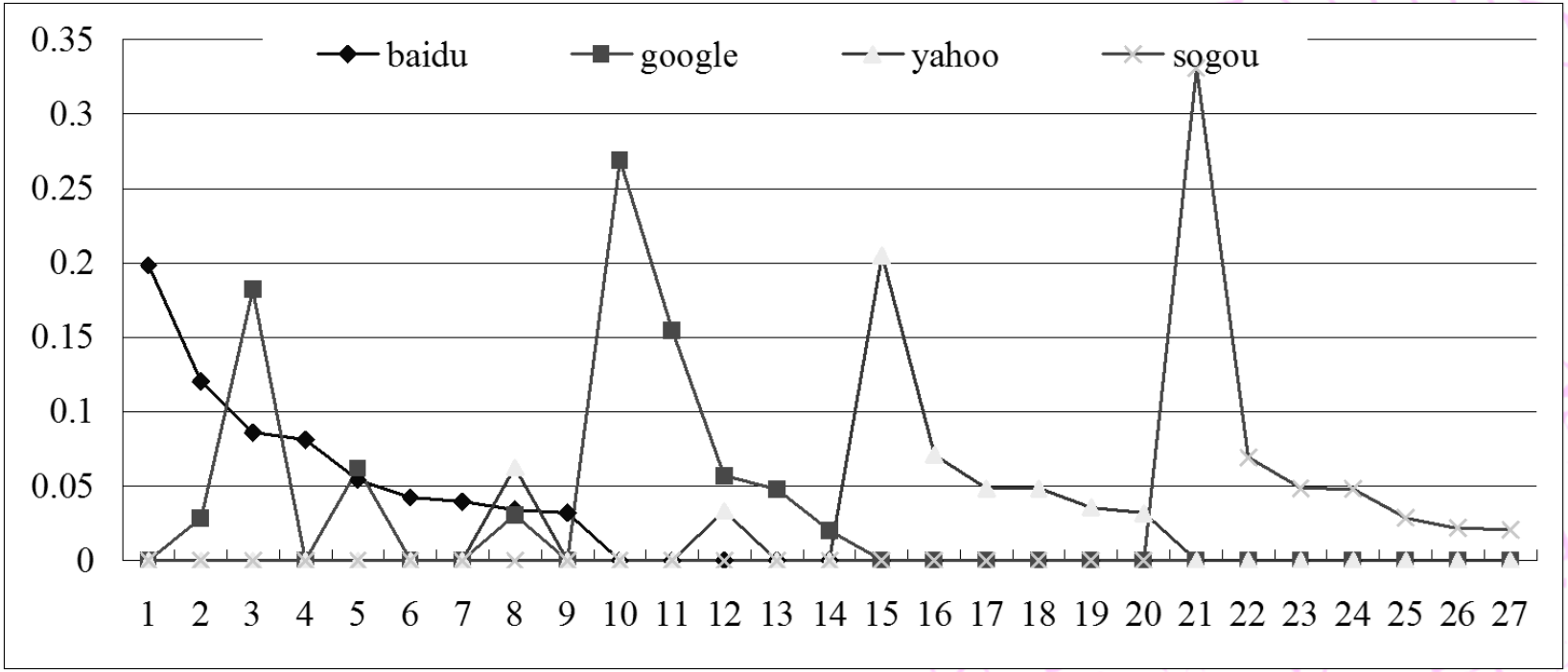
Automatic answer annotation

* For navigational type queries (e.g. yahoo mail)
  * Basic assumption: *The result clicked by more users should be more relevant than the one clicked by fewer users.*



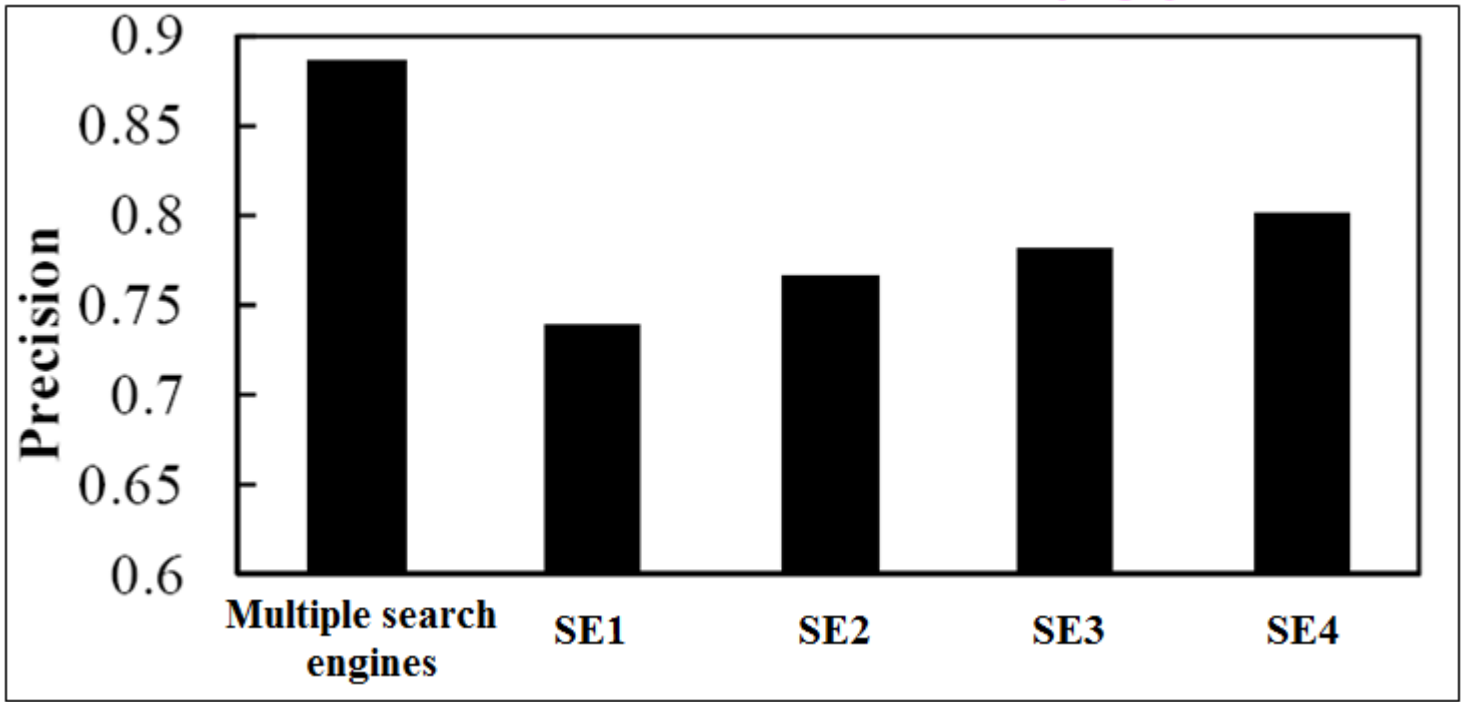
Automatic answer annotation

* For informational/transactional type queries
  * The basic assumption fails for non-navigational type queries. e.g. the query “电影” (movie)



Automatic answer annotation

* For informational/transactional type queries (cont.)
  * Improved assumption: *click-through data from multiple search engines are more informational and less biased than that from a single engine*.



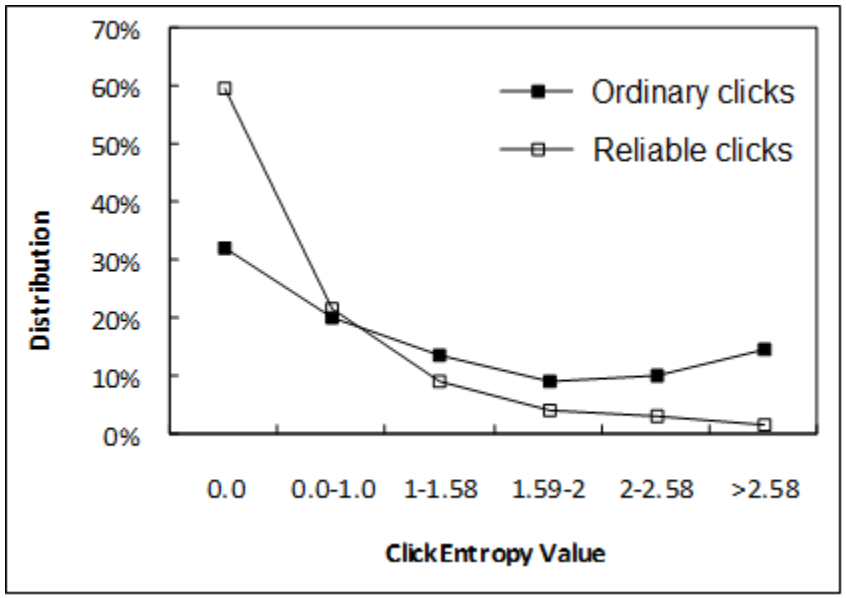
Automatic answer annotation

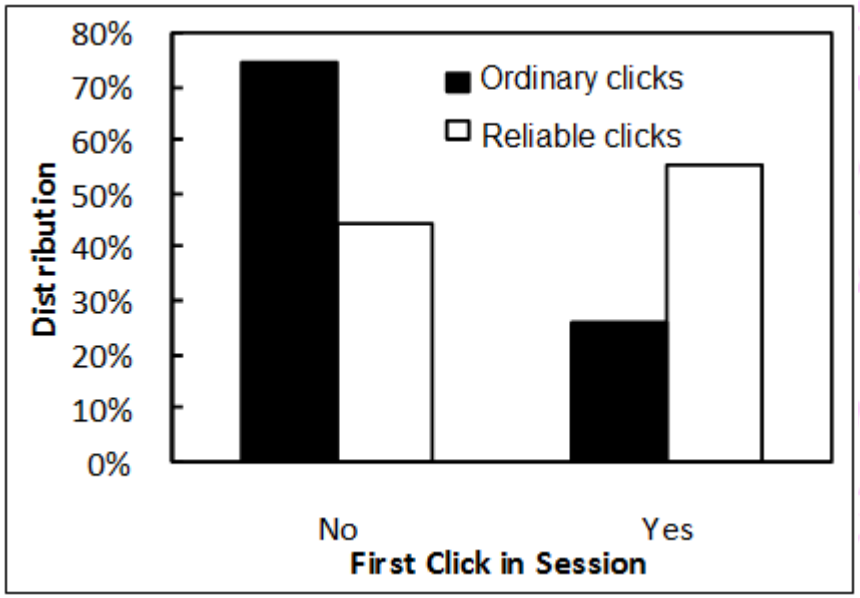
* For long-tail queries (cont.)
  * Only a few clicks for a long-tail query: each click should make difference in answer annotation process.

No.	Feature Name	Description
1	NumQueries	Num of unique queries in current session
2	ClickEntropy	Entropy of click distribution in current session
3	ClickSelfInformation	Self-information of current doc in current session
4	ClickDocNumInSession	Click number of current doc in current session
5	ClickDocNumInQuery	Click number of current doc in current query
6	IsFirstClickInSession	=1 if the first click in current session, =0 otherwise
7	IsLastClickInSession	=1 if the last click in current session,=0 otherwise
8	IsFirstClickInQuery	=1 if the first click in current query, =0 otherwise
9	IsLastClickInQuery	=1 if the last click in current query, =0 otherwise
10	ClickOrderInSession	Order of current click in current session
11	ClickOrderInQuery	Order of current click in current query
12	ClickDocRank	Result rank of current click

Automatic answer annotation

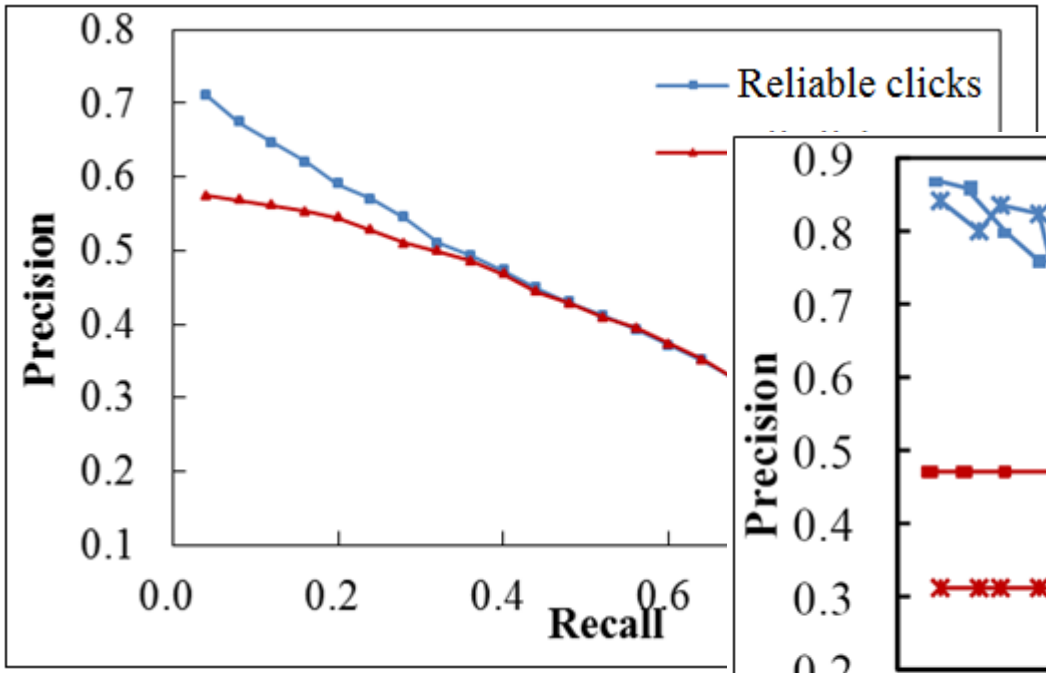
- For long-tail queries (cont.)
  - Clicks with reliable relevance feedback information is different from unreliable ones.
  - A learning based framework can be adopted to separate reliable clicks.





Automatic answer annotation

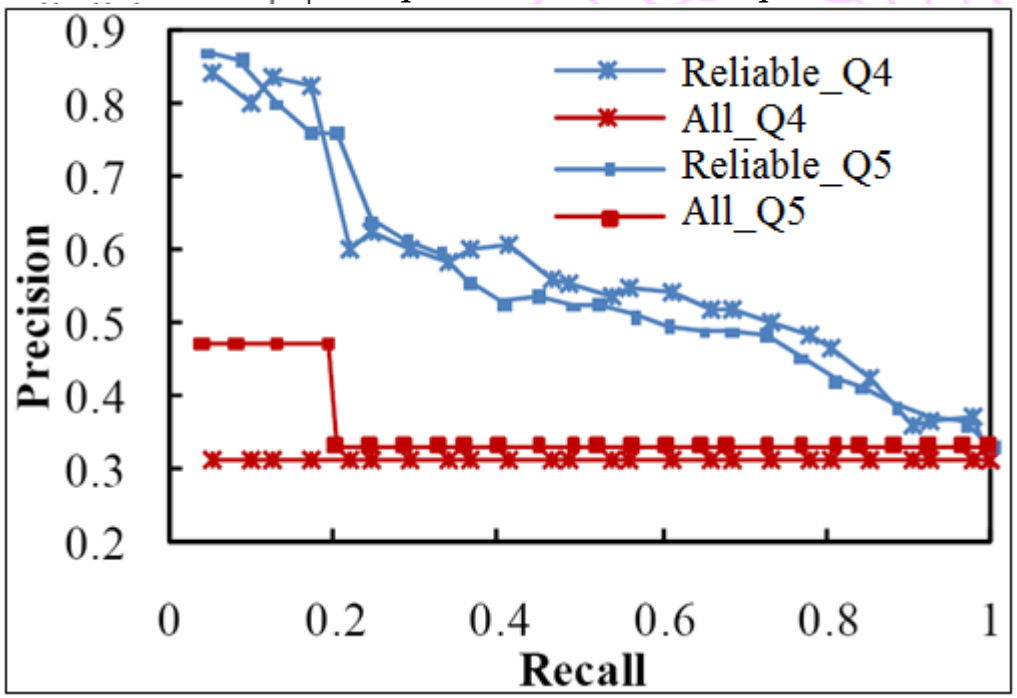
* For long-tail queries (cont.)
  * Evaluation results



Q1: queries with at least 100 requests

Q4: queries with only one user request

Q5: queries with less than or equal to two user requests



User satisfaction evaluation

* Problems with Cranfield-like approaches
  * Time consuming, objective
  * Relevance annotation of "query-result" pairs
  * Ignore the representation of results

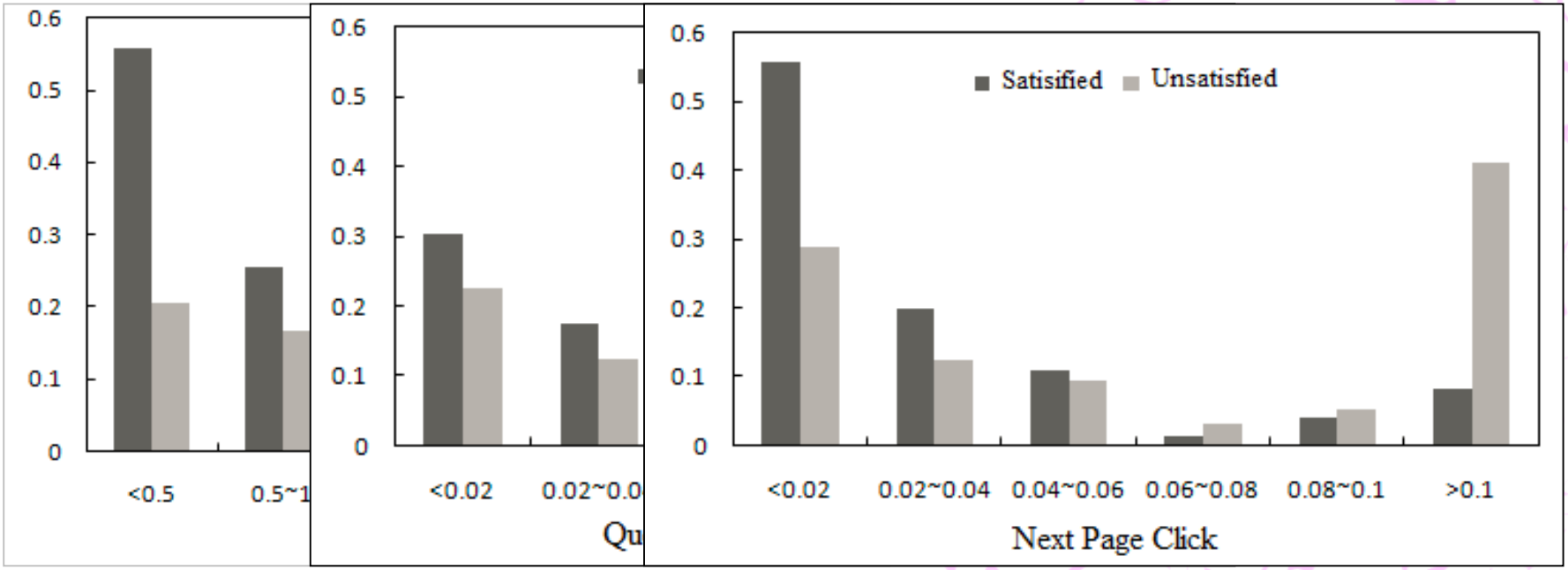
nks

User satisfaction evaluation

* User satisfaction evaluation instead of relevance judgment
  * What's a satisfied user session?
  * Navigational: top result should be the target.
  * Informational: top ranked results answer user's question with a different aspect.
  * Transactional: user can accomplish task with the top few results.
  * Behavior patterns in satisfied/unsatisfied search sessions should be different.

User satisfaction evaluation

* A number of behavior features
  * Result click behavior: first click position, last click position, revisit click, non-click, ...
  * Other click behavior: recommendation click, next page



User satisfaction evaluation

* Compared with human assessors

	AUC	AUC difference
Human assessor	0.87	/
Informational/Transactional	0.75	-16.00%
Navigational	0.80	-8.75%

	A	B	System
Assessor A as ground truth	1.00	0.80	0.80
Assessor B as ground truth	0.80	1.00	0.76
System as ground truth	0.80	0.76	1.00

* Query frequency v.s. applicable queries

Query frequency	1	2~3	4~10	11~100	100~	top
Percentage of Applicable queries	2.59%	14.96%	36.11%	65.61%	89.56%	94.11%

Search performance evaluation

* Selected publications
  * Bo Zhou, Yiqun Liu, Min Zhang, Yijiang Jin, Shaoping Ma, Incorporating Web Browsing Information into Anchor Texts for Web Search, **Information Retrieval.** Volume 14, Issue 3: 290-314, 2011.
  * Danqing Xu, Yiqun Liu, Min Zhang and Shaoping Ma. Incorporating Revisiting Behaviors into Click Models. Accepted by **the 5th ACM International Conference on Web Search and Data Mining** . WSDM 2012.
  * Yiqun Liu, Yupeng Fu, Min Zhang, Shaoping Ma, Liyun Ru. Automatic Search Engine Performance Evaluation with Click-through Data Analysis. in Proceedings of the 16th international Conference on World Wide **WWW '07**. ACM, New York, 1133-1134.
  * Rongwei Cen, Yiqun Liu, Min Zhang, Bo Zhou, Liyun Ru, Shaoping Ma. Exploring Relevance for Clicks. In Proceeding of the 18th ACM Conference on information and Knowledge Management. **CIKM '09**. 1847-1850.

Conclusions

* Key challenges of search engines
* Meet challenges with the help of "wisdom of the crowd"
  * User behavior and information need
  * Web spam fighting
  * Search performance evaluation

Thank you

Welcome to visit our homepage

http://www.thuir.cn/

On-line Demos

Search Engine Evaluation

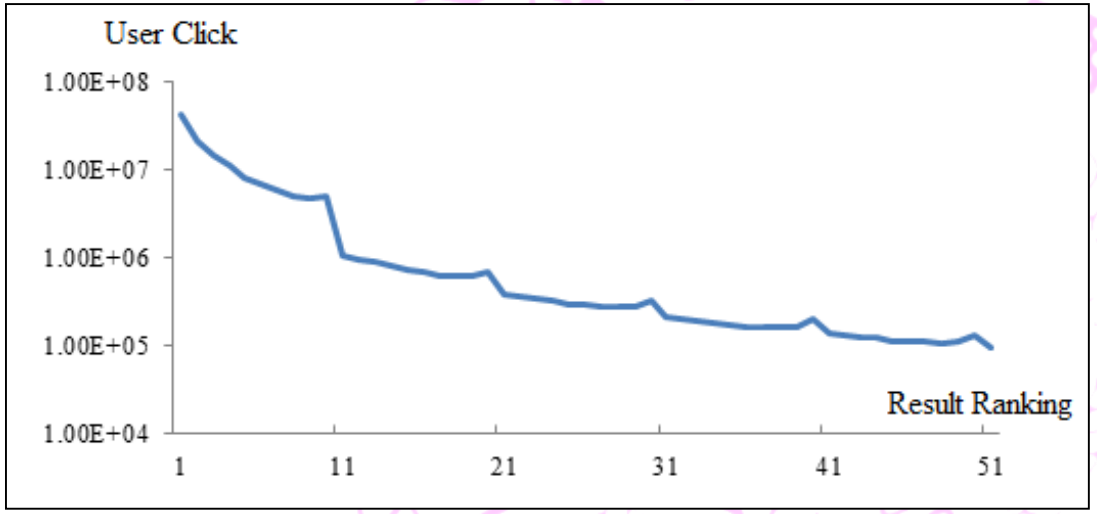
Seasonal Epidemic Prediction

Web Spam Page Identification

Web News Event Clustering

Click model construction

- Problems with the automatic answer annotation process
  - Each click is regarded as a relevance voting for the corresponding result.
  - However, results aren't equally examined: **Position Bias**



Click model construction

- Model click behavior to solve the position bias problem

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

- How to estimate the examine probability?
  - Cascade model: $P(E_{i+1} = 1 | E_i = 1, C_i) = 1 - C_i$
  - Dependent click model (DCM):

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

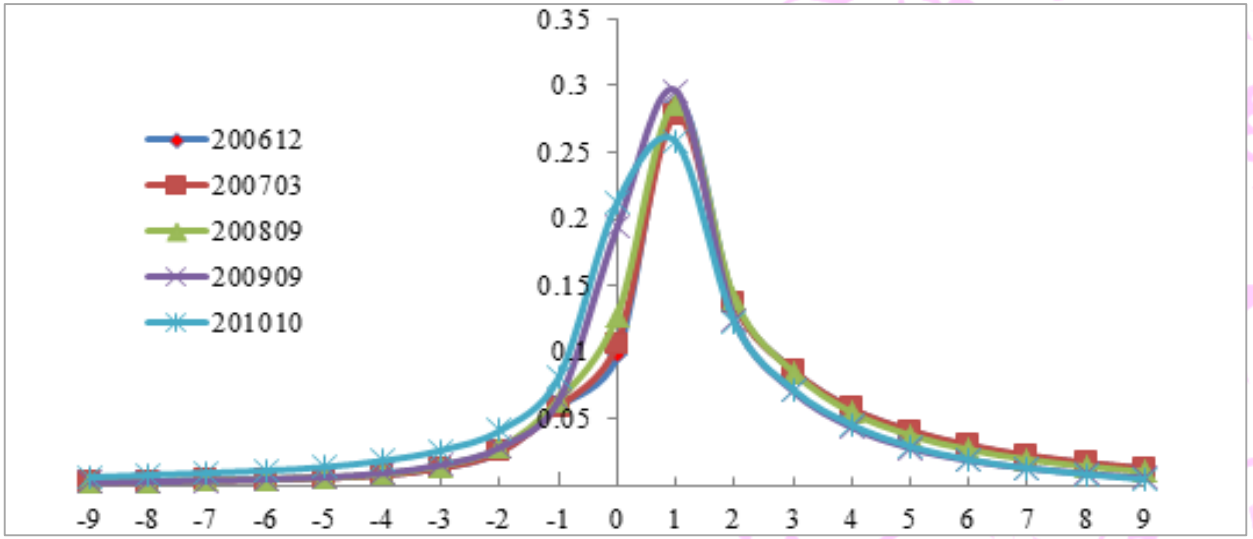
  - User browsing model (UBM):

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

  - Lots of other models: DBM, CCM, ...

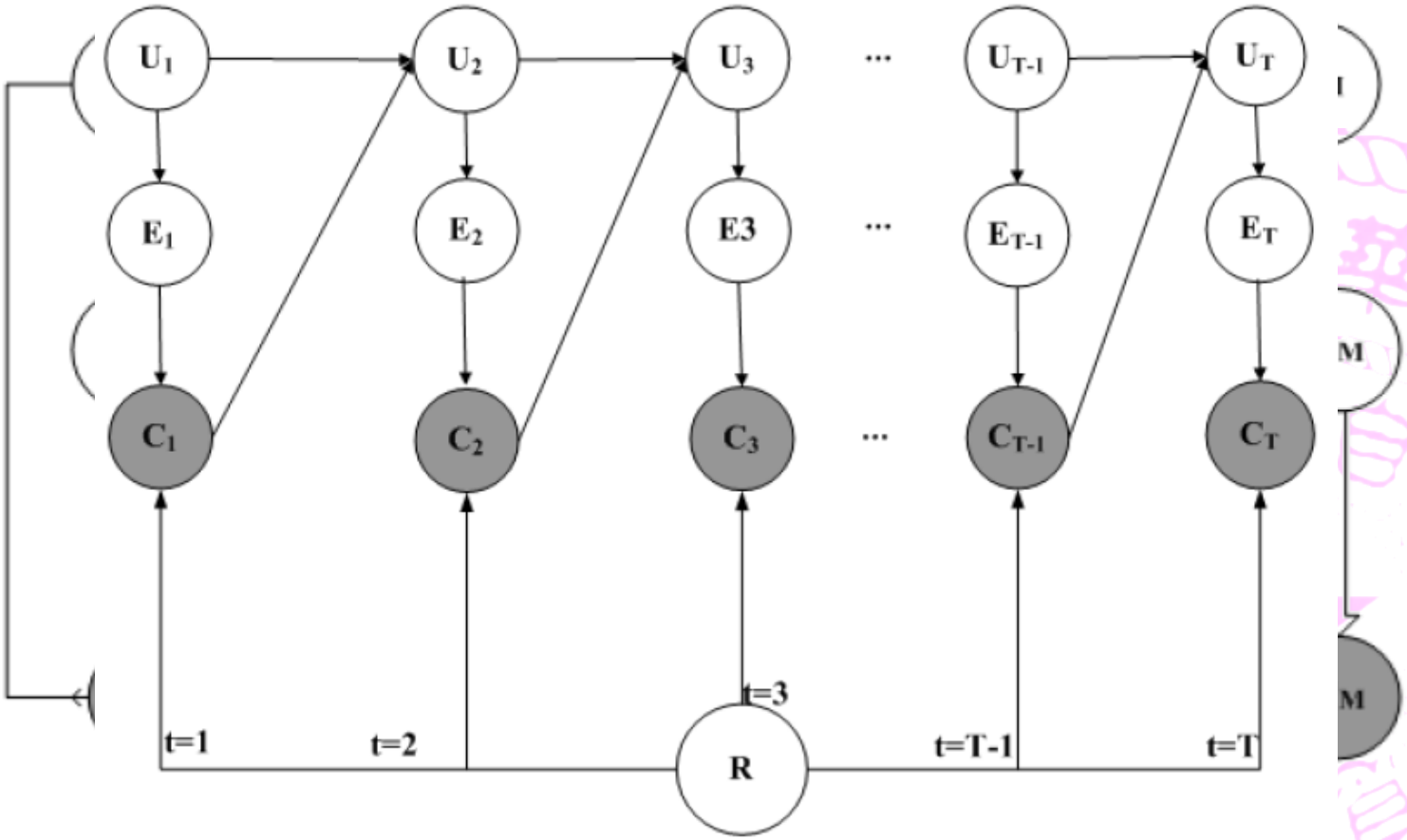
Click model construction

* Problem with the existed models
  * Results are not sequentially examined and clicked
  * Eye-tracking experiments (Lorigo et.al, 2005) show that lots of users revisit
  * Revisit behavior is popular (24.1% multi-click sessions)



Click model construction

* From ranking position to timing sequence



Click model construction

* The Temporal Hidden Click Model (THCM)

$$p(E_{t(i+1)} = 1 | E_{ti} = 1) = \alpha$$ Forward examination

$$p(E_{t(i-1)} = 1 | E_{ti} = 1) = \gamma$$ Backward examination (revisit)

$$0 \leq \alpha + \gamma \leq 1, \alpha \geq 0, \gamma \geq 0$$

$$p(C_t | C_1, C_2, ..., C_{t-1}) = p(C_t | C_{t-1})$$ One-order model (to be simple)

$$p(C_t = i) = p(E_{ti} = 1).p(R_i = 1)$$ similar with old models

Data requirement: the click sequence should be recorded

Click model construction

* THCM performance