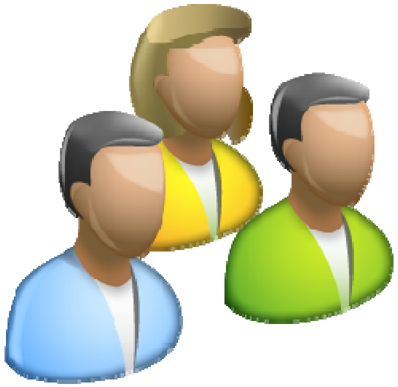
More Than Relevance: High Utility Query Recommendation By Mining Users' Search Behaviors

Xiaofei Zhu, **Jiafeng Guo**, Xueqi Cheng, Yanyan Lan
Institute of Computing Technology, CAS

Motivation

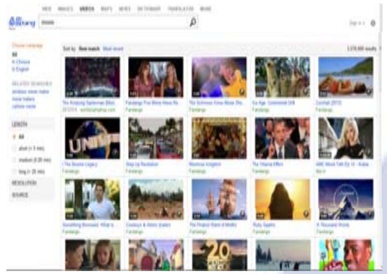
Information Seeking Tasks



Query

not easy to formulate properly

Find Web pages



Locate resources

Access Info of topics

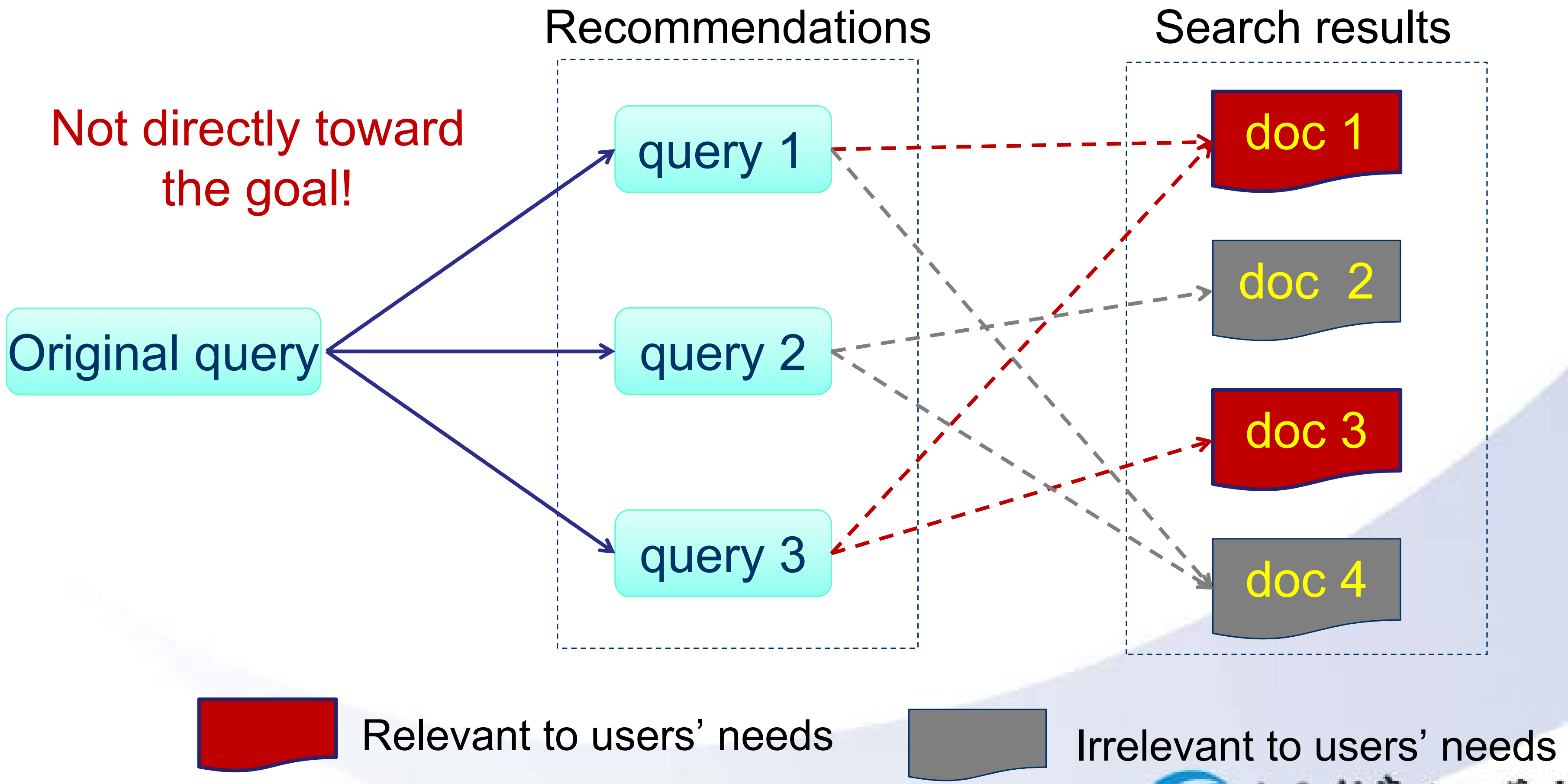
The ultimate goal of query recommendation

Assist users to reformulate queries so that they can **acquire their desired information successfully and quickly**

Motivation

Relevant query recommendation:
Providing alternative queries similar to a user's initial query

Problem:
relevant query ~~→~~ satisfy users' needs
not necessarily

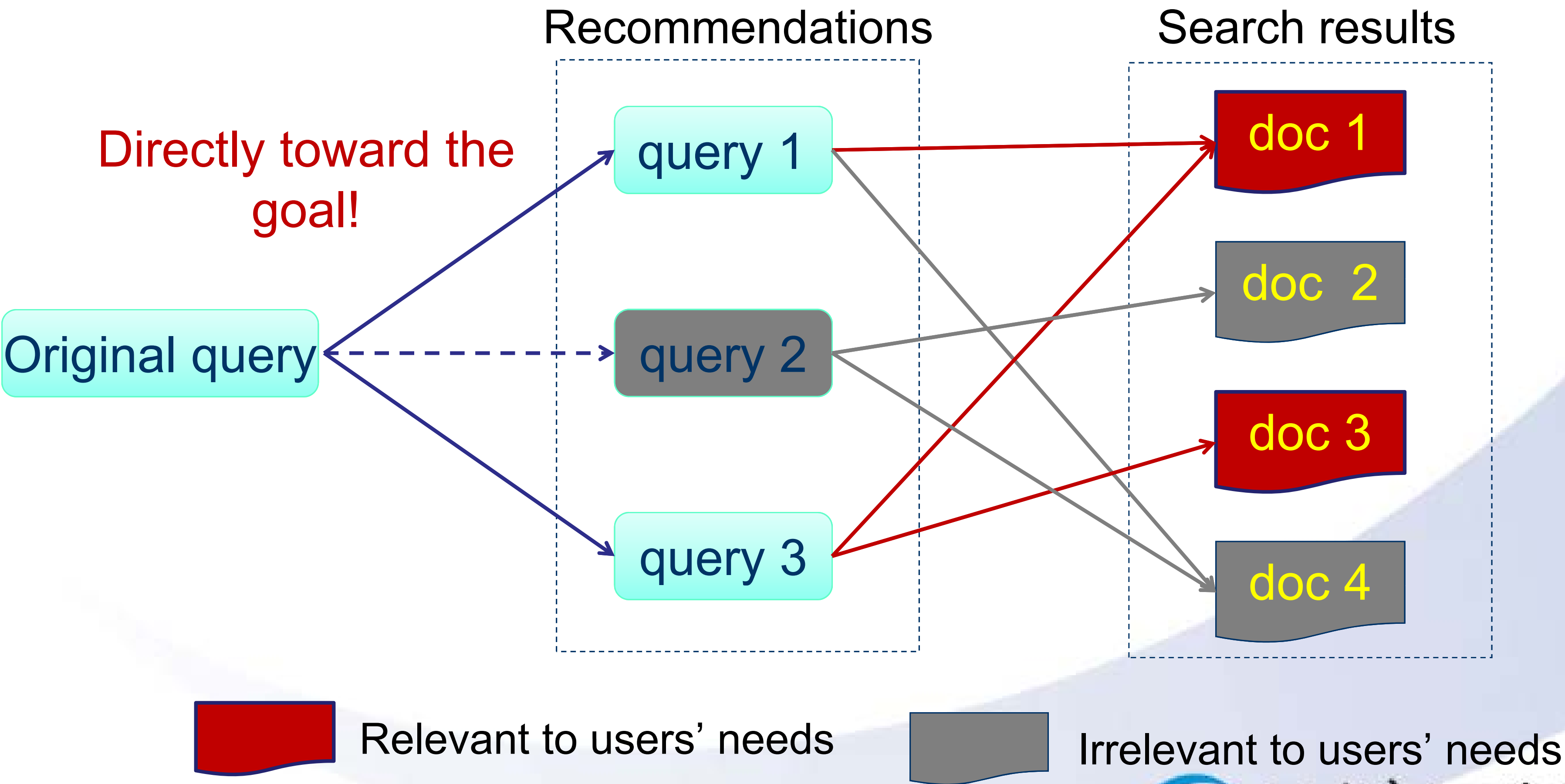


Motivation

High Utility Recommendation:
Providing queries that can better satisfy users' information needs

Query Utility Definition:

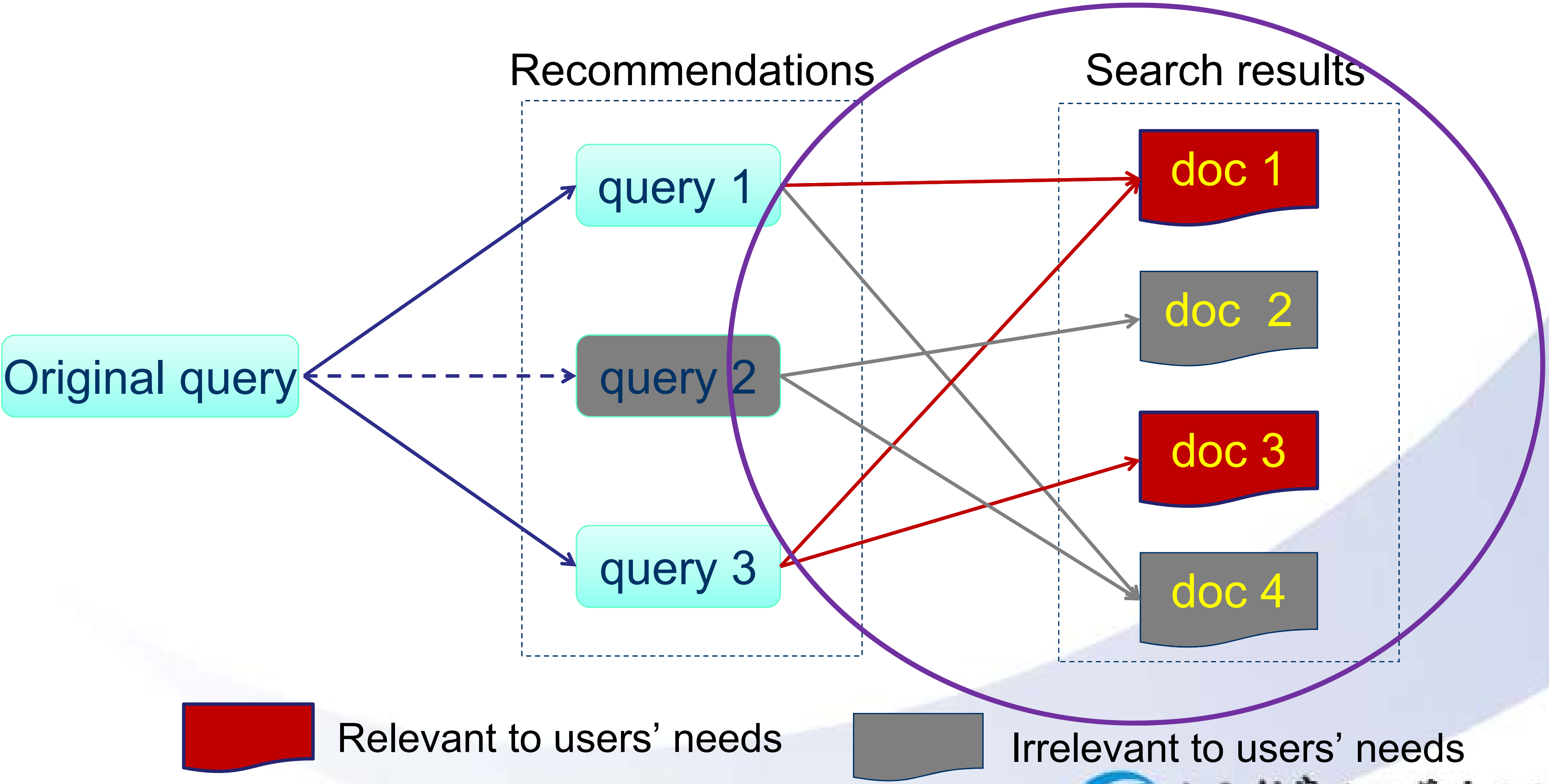
The information gain that a user can obtain from the search results of the query according to her original information needs.



Motivation

true effectiveness of query recommendation

High Utility Recommendation ⇨ Emphasize users' post-click satisfaction



Challenges for high utility recommendation

→ How to infer query utility?

Query Utility Model

→ How to evaluate?

Two evaluation metrics

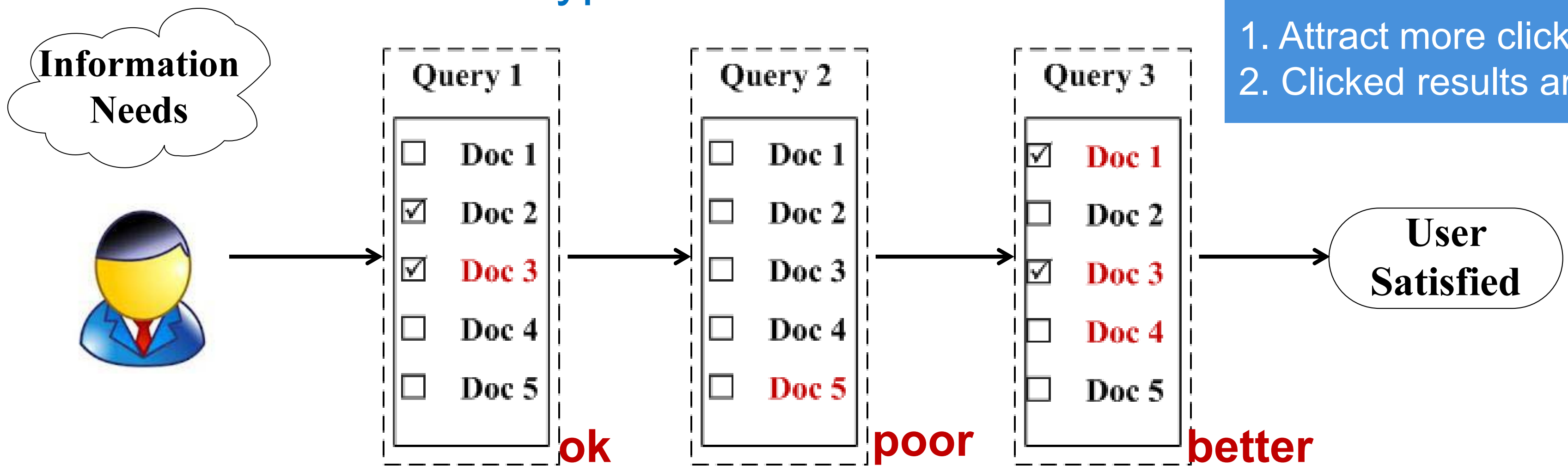
Our Approach

how to infer query utility ?

Key Idea: Through user's search behaviors

A typical search session

1. Attract more clicks
2. Clicked results are relevant



Red - relevant √ - clicked

Perceived Utility

model the attractiveness of the search results

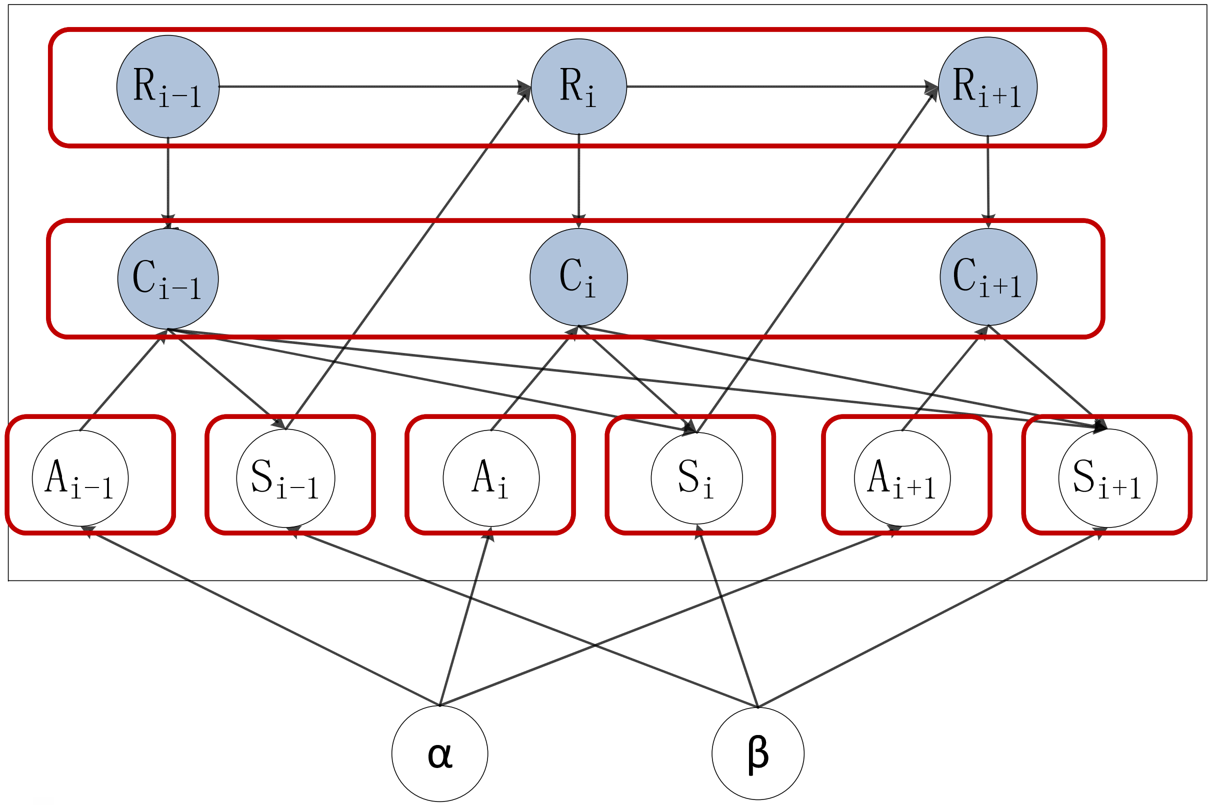
Query Utility

Posterior Utility

model the satisfaction of the clicked search results

Query Utility Model (dynamic Bayesian network)

how to infer query utility ?



$$P(R_i = 1 | R_{i-1} = 1, S_{i-1} = 1) = 0.$$

$$P(C_i = 1 | R_i = 1, A_i = 1) = 1,$$

$$P(A_i = 1) = \alpha_{\phi(i)},$$

$$P(S_i = 1 | C_{1:i}) = \sigma(\sum_{k=1} \beta_{\phi(k)} \cdot I(C_k = 1)),$$

$$\sigma(x) = \frac{1}{1+e^{-x}}.$$

Perceived Utility α : control the probability of the attractiveness

Posterior Utility β : control the probability of users' satisfaction

R_i : whether there is a reformulation at position i

C_i : whether the user clicks on sor… …ation at position i;

A_i : whether the user is attracted by the search results of the reformulation at position I;

Query Utility $\mu_t = \alpha_t * \beta_t$

The expected information gain users obtained from the search results of the query according to their original information needs

Parameter Estimation

how to infer query utility ?

Maximum Likelihood Estimation

$$P(C_{1:M}, R_{1:M}, A_{1:M}, S_{1:M})$$

$$= \prod_{i=1}^{M} P(C_i \mid A_i, R_i) \cdot (R_i \mid R_{i-1}, S_{i-1}) \cdot P(A_i) \cdot P(S_i \mid C_{1:i})$$

$$\alpha_t = \frac{\sum_{j=1}^{N} \sum_{i=1}^{M} A_i^j \cdot I(\phi_j(i) = t)}{\sum_{j=1}^{N} \sum_{i=1}^{M} I(\phi_j(i) = t)}$$

$$= \frac{\sum_{j=1}^{N} \sum_{i=1}^{M} I(C_i^j = 1) \cdot I(\phi_j(i) = t)}{\sum_{j=1}^{N} \sum_{i=1}^{M} I(\phi_j(i) = t)}$$

$$\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta_t^2} - 2\mu & 0 & 2\lambda_t \\ -1 & 2z_t & 0 \\ \lambda_t^2 & 0 & 2\lambda_t \beta_t \end{bmatrix} \begin{bmatrix} \Delta\beta_t \\ \Delta z_t \\ \Delta\lambda_t \end{bmatrix} = - \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \beta_t} - 2\mu\beta_t + \lambda_t^2 \\ -\beta_t + z_t^2 \\ \lambda_t^2 \beta_t \end{bmatrix}$$

$$\begin{cases} \Delta\beta_t = \dfrac{-\beta_t \frac{\partial \mathcal{L}}{\partial \beta_t} + 2\mu\beta_t^2}{\beta_t \frac{\partial^2 \mathcal{L}}{\partial \beta_t^2} - 2\mu\beta_t - \lambda_t^2}, \\ \Delta z_t = \dfrac{\beta_t - z_t^2 + \Delta\beta_t}{2z_t}, \\ \Delta\lambda_t = \dfrac{-\lambda_t^2\beta_t - \lambda_t^2\Delta\beta_t}{2\lambda_t\beta_t}. \end{cases}$$

Newton-Raphson Algorithm

Evaluation

how to evaluate ?

Query Level Judgment

Original query

Recommendations

query 1	Relevant or Not?
query 2	Relevant or Not?
query 3	Relevant or Not?

Relevant = 1

Partial Relevant = 0.5

Irrelevant = 0

Evaluation

how to evaluate ?

Document Level Judgment

Original query

Recommendations & Clickthrough

query 1

doc 1 doc 2 doc 3

Relevant or Not?

query 2

doc 1

Relevant or Not?

query 3

doc 1 doc 2

Relevant or Not?

Evaluation

how to evaluate ?

- QRR (Query Relevant Ratio)

$$QRR(q) = \frac{RQ(q)}{N(q)}$$

Measuring the probability that a user finds(clicks) relevant results when she uses query q for her search task.

- MRD (Mean Relevant Document)

$$MRD(q) = \frac{RD(q)}{N(q)}$$

Measuring the average number of relevant results a user finds(clicks) when she uses query q for her search task.

Experiments

- Dataset: UFindIt log data (SIGIR'11 Best Paper)
  - A period of 6 months, consisting 1484 search sessions conducted by 159 users (reformulation and click).
  - Manual relevant judgments on results with respect to the original needs

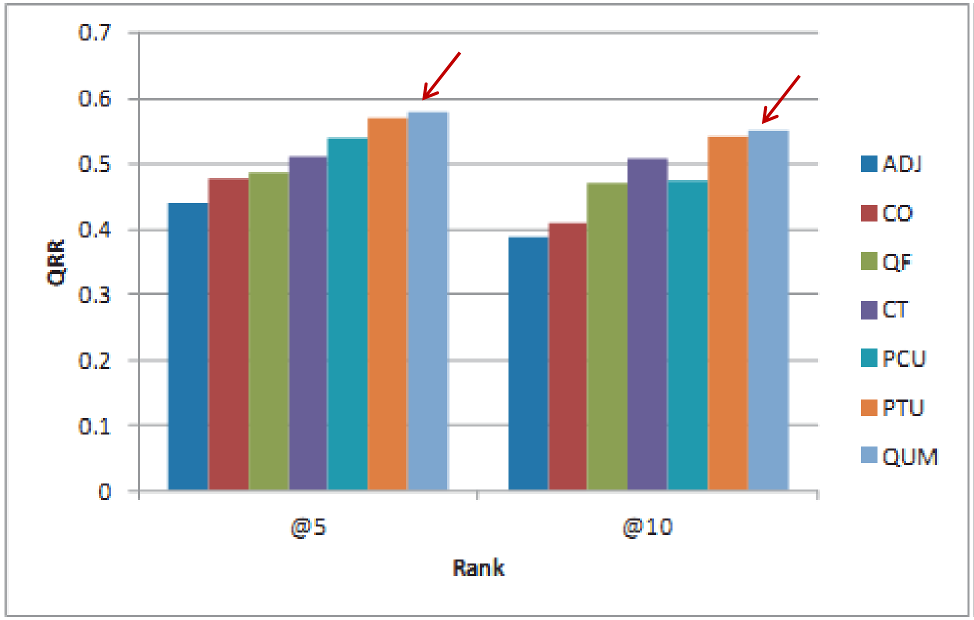
- Data Processing:
  - We process the data by ignoring some interleaved sessions, remove sessions which have no reformulations, and sessions started without queries, after processing, we obtain:
    - 1,298 search sessions
    - 1,086 distinct queries
    - 1,555 distinct clicked URLs
  - For each test query, the average number of search sessions is 32 and the average number of distinct candidate queries is 26.

Baseline Methods

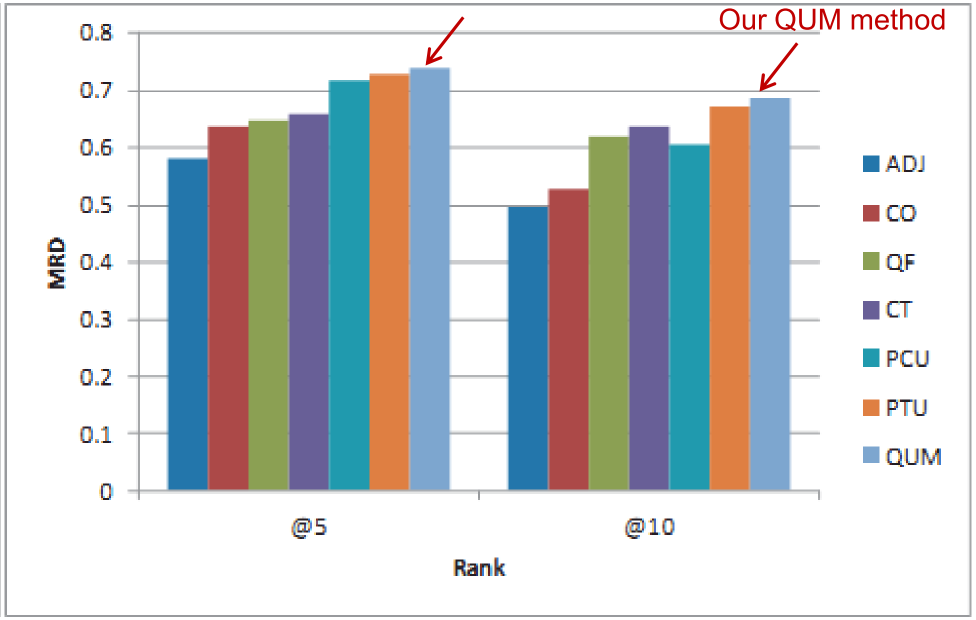
- Frequency-based methods
  - Adjacency (ADJ) (WWW 06)
  - Co-occurrence (CO) (JASIST 03)
- Graph-based methods
  - Query-Flow Graph (QF) (CIKM 08)
  - Click-through Graph (CT) (CIKM 08)
- Component utility methods
  - Perceived Utility (PCU)
  - Posterior Utility (PTU)

Experimental Results

Comparison of the performance of all approaches (ADJ,CO,QF,CT,PCU,PTU,QUM) in terms of QRR and MRD.



(a) QRR



(b) MRD

The performance improvements are significant (t-test，p-value <= 0.05)

Experimental Results

The improvement is larger on difficult queries!

Query Difficulty	Method	QRR		MRD	
		@5	@10	@5	@10
Easy	ADJ	0.588(18.64%)	0.526(26.30%)	0.771(20.32%)	0.674(25.22%)
	CO	0.609(14.55%)	0.529(25.63%)	0.830(11.80%)	0.687(22.89%)
	QF	0.618(12.94%)	0.604(9.89%)	0.846(9.67%)	0.806(4.69%)
	CT	0.654(6.62%)	0.635(4.65%)	0.836(11.02%)	0.805(4.79%)
	PCU	0.656(6.37%)	0.611(8.74%)	0.889(4.35%)	0.798(5.79%)
	PTU	0.689(1.22%)	0.663(0.17%)	0.908(2.18%)	0.837(0.86%)
	QUM	**0.698**	**0.664**	**0.928**	**0.844**
Medium	ADJ	0.460(30.00%)	0.429(33.19%)	0.596(24.14%)	0.527(33.76%)
	CO	0.495(20.81%)	0.441(29.65%)	0.640(15.72%)	0.550(28.10%)
	QF	0.511(17.07%)	0.500(14.39%)	0.615(20.43%)	0.630(11.79%)
	CT	0.534(12.07%)	0.549(4.02%)	0.689(7.54%)	0.692(1.81%)
	PCU	0.544(9.91%)	0.485(17.74%)	0.703(5.31%)	0.588(19.76%)
	PTU	0.581(2.87%)	0.557(2.70%)	0.722(2.53%)	0.689(2.18%)
	QUM	**0.598**	**0.572**	**0.740**	**0.704**
Hard	ADJ	0.259(65.27%)	0.216(91.19%)	0.351(54.37%)	0.284(77.27%)
	CO	0.314(36.29%)	0.261(58.17%)	0.412(31.63%)	0.340(48.00%)
	QF	0.324(32.08%)	0.312(32.20%)	0.441(22.94%)	0.414(21.78%)
	CT	0.334(28.08%)	0.343(20.17%)	0.437(24.15%)	0.424(18.85%)
	PCU	0.404(5.90%)	0.324(27.07%)	0.534(1.54%)	0.413(22.02%)
	PTU	0.426(0.28%)	0.402(2.51%)	0.526(3.18%)	0.485(3.92%)
	QUM	**0.427**	**0.412**	**0.542**	**0.504**

Conclusions

- Contribution
  - Recommend high utility queries rather than only relevant queries: to directly toward the ultimate goal of query recommendation;
  - A novel dynamic Bayesian network (i.e., QUM) to mine query utility from users' reformulation and click behaviors;
  - Introduce two evaluation metrics for utility based recommendation
  - Evaluate the performance on a real query log and show the effectiveness

- Future work
  - Extend our utility model to capture the specific clicked URLs for finer modeling

Thanks!

guojiafeng@ict.ac.cn