Information Retrieval 1

IR-User Interaction

Ilya Markov
i.markov@uva.nl

University of Amsterdam

Learning to rank

Evaluation

Document representation & matching

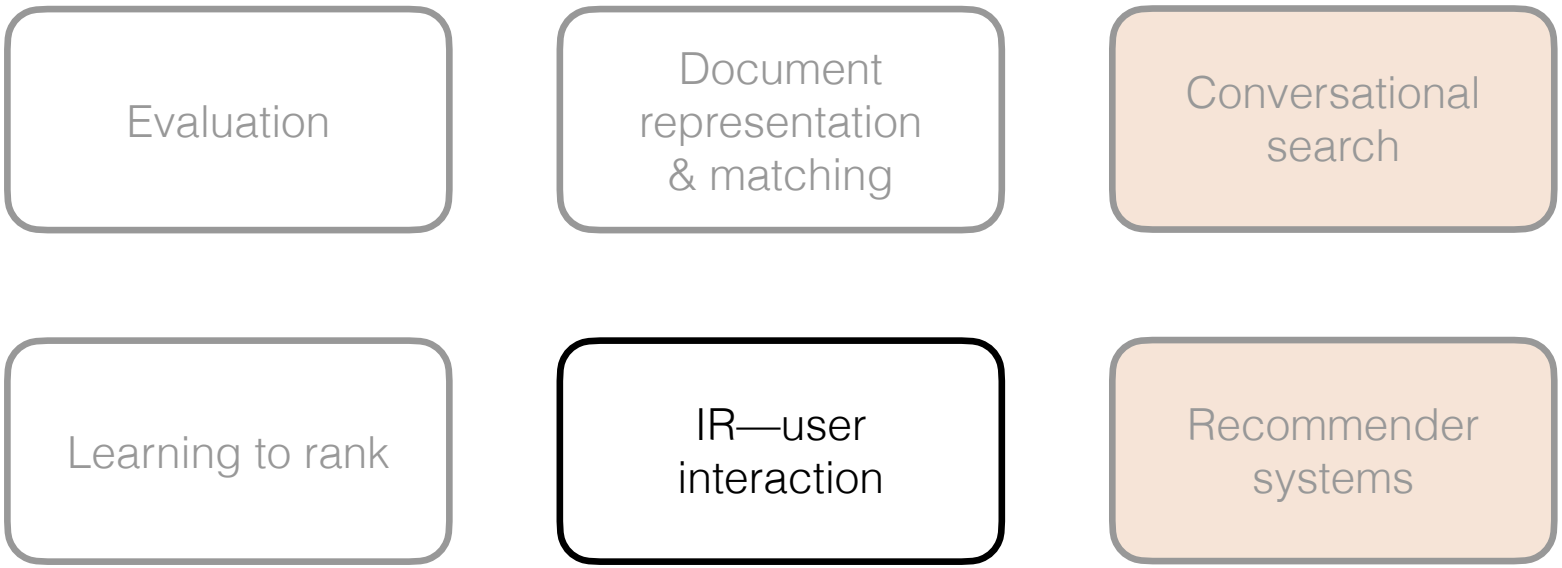
Conversational search

Learning to rank

IR—user interaction

Recommender systems

IR-User Interaction



User interactions

- Queries
- Interactions with a SERP (clicks, mousing, scrolling, etc.)
- Time between user actions
- Closing browser
- Interactions beyond search
- Etc.

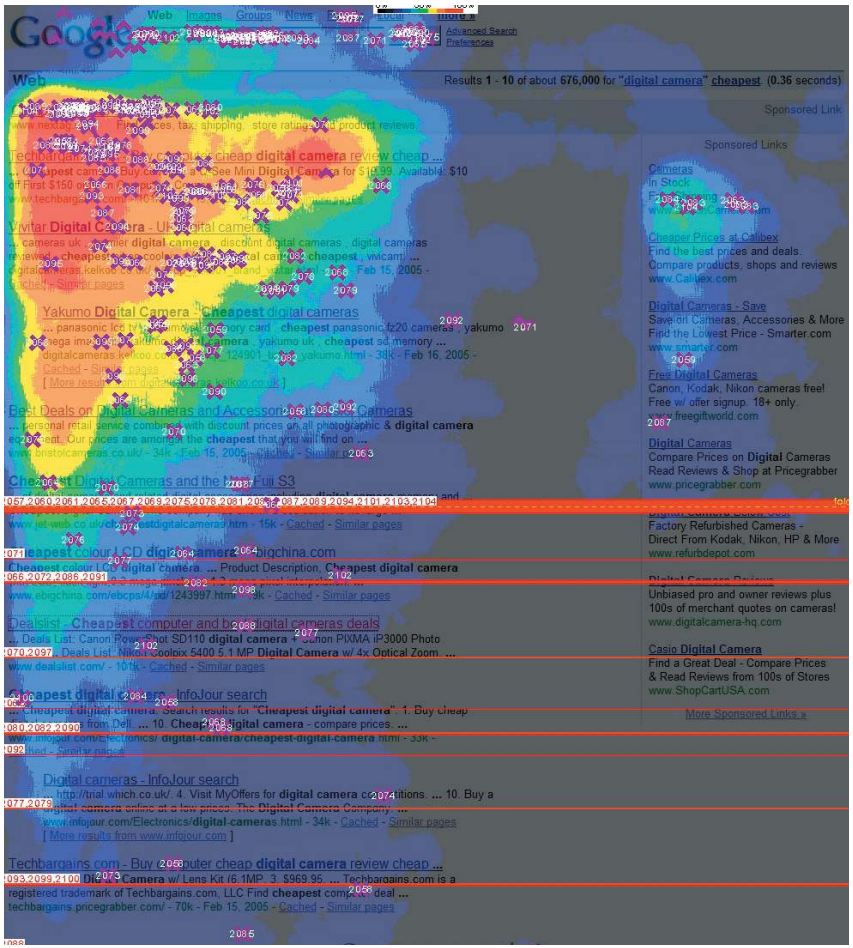
Why are user interactions important?

- Evaluate IR systems
- Improve IR systems

Models of user search interactions

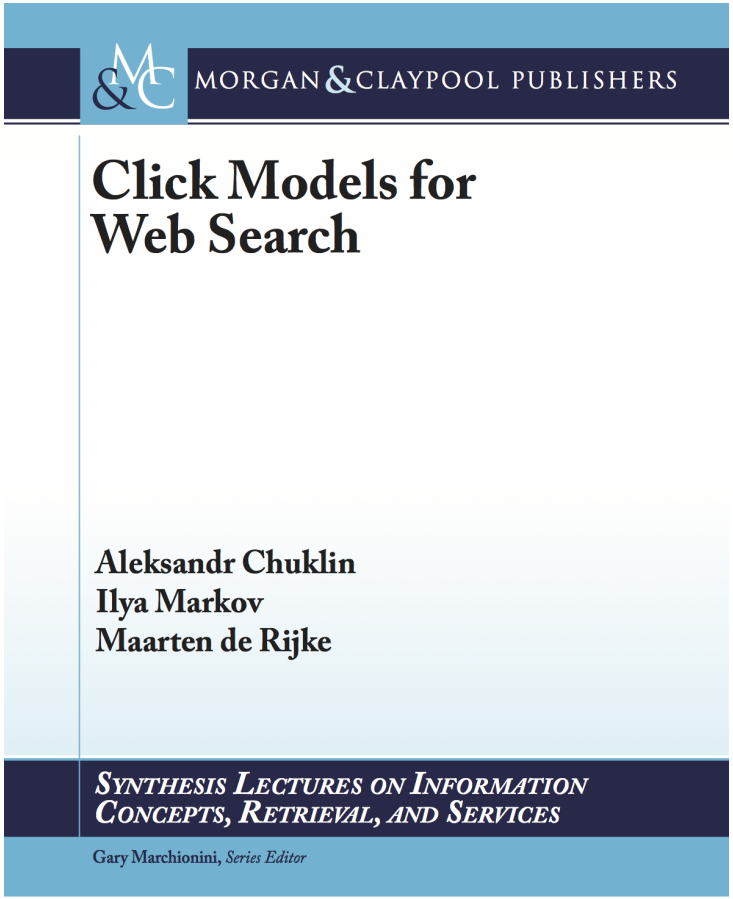
- Click models
- Models of mouse hovering
- Models of time between user actions

Position bias



Hotchkiss et al. "An In Depth Look at Interactions with Google using Eye Tracking Methodology"

Click models



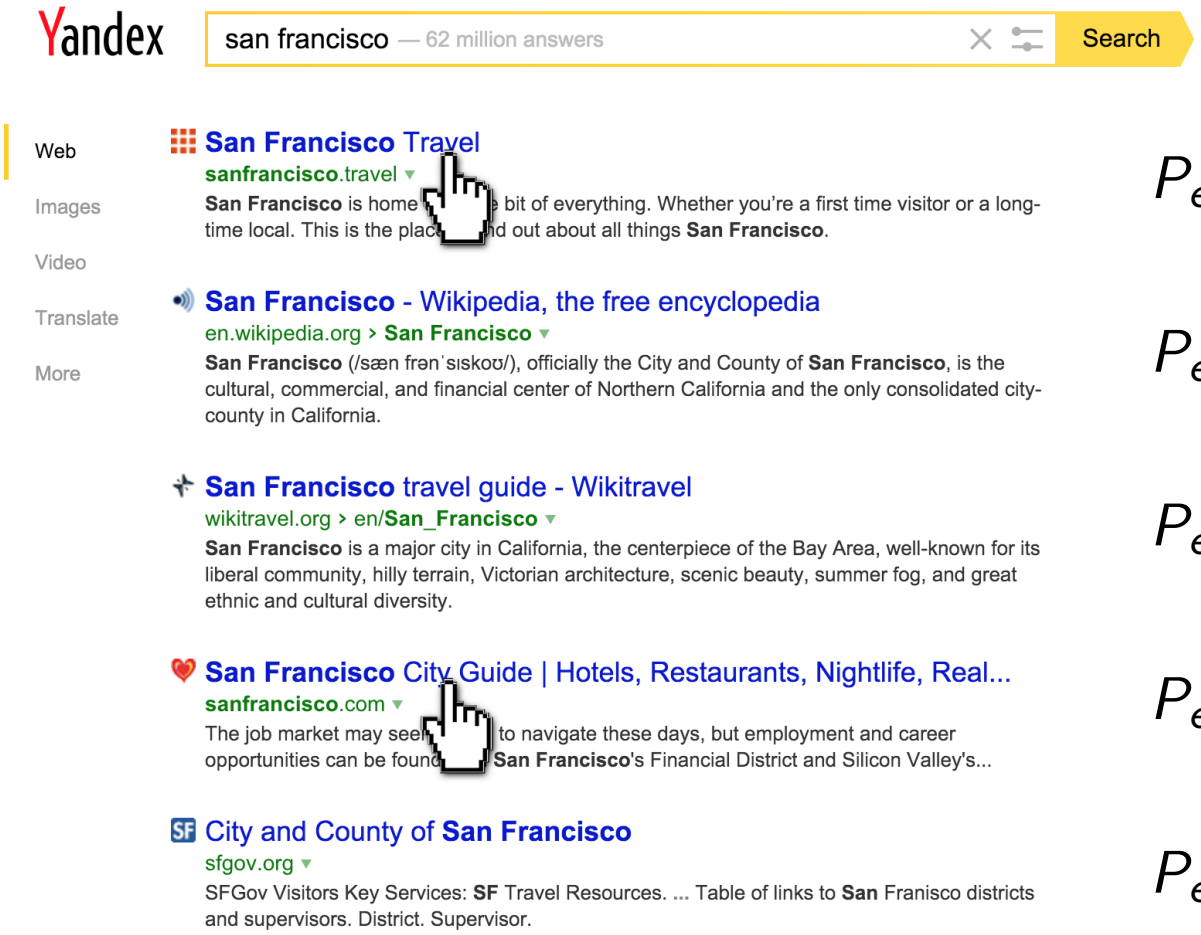
http://clickmodels.weebly.com/the-book.html

Outline

Basic click models

- Position-based model
- Cascade model

Position-based model



$P_{exam}(1),\ P_{attr}(qd_1)$

$P_{exam}(2),\ P_{attr}(qd_2)$

$P_{exam}(3),\ P_{attr}(qd_3)$

$P_{exam}(4),\ P_{attr}(qd_4)$

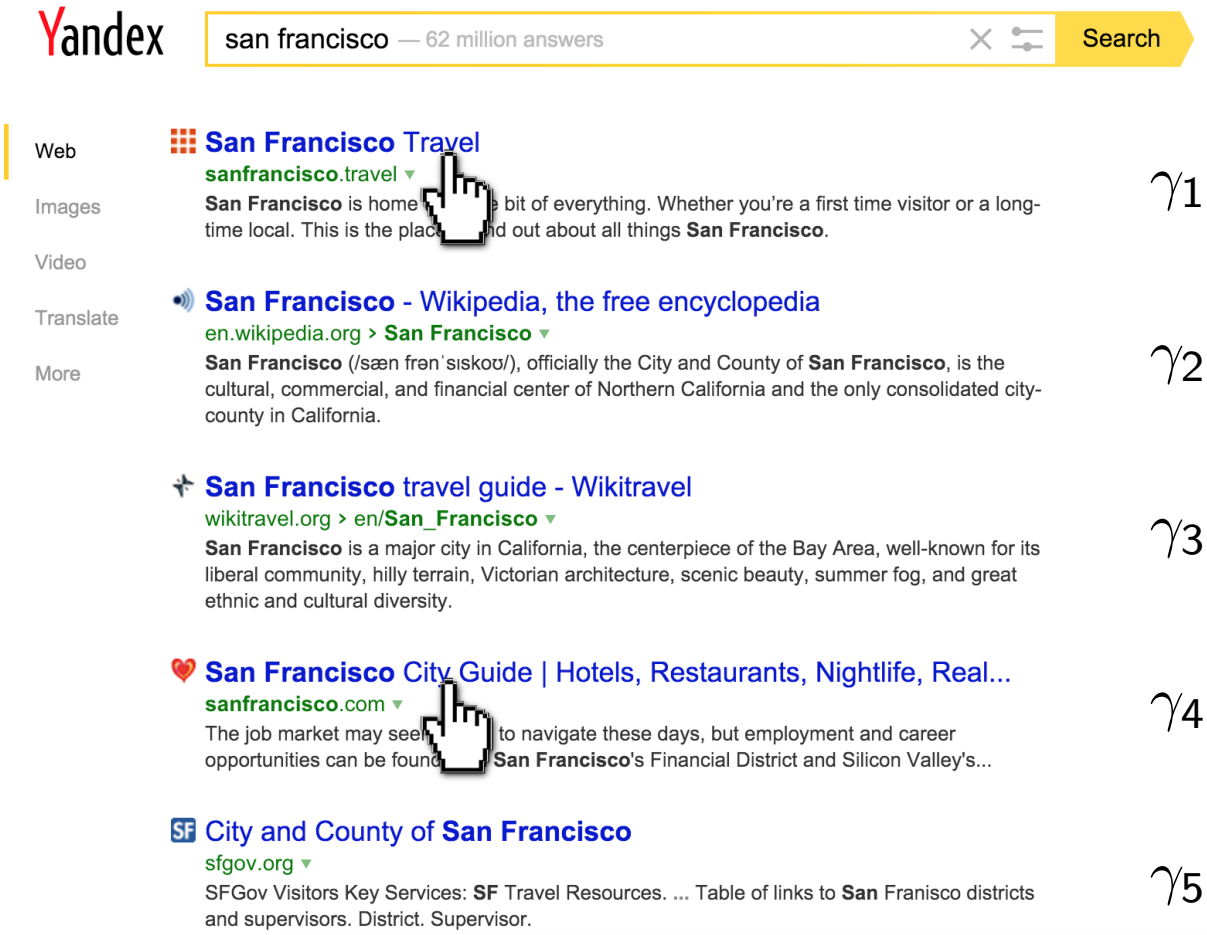
$P_{exam}(5),\ P_{attr}(qd_5)$

Position-based model: examination

- Terminology
  - Examination = reading a **snippet**
  - E_r – binary random variable denoting examination of a snippet at rank r
- Position-based model (PBM)
  - Examination depends on rank

$$P(E_r = 1) = \gamma_r$$

Position-based model



γ_1 , $P_{attr}(d_1 q)$

γ_2 , $P_{attr}(d_2 q)$

γ_3 , $P_{attr}(d_3 q)$

γ_4 , $P_{attr}(d_4 q)$

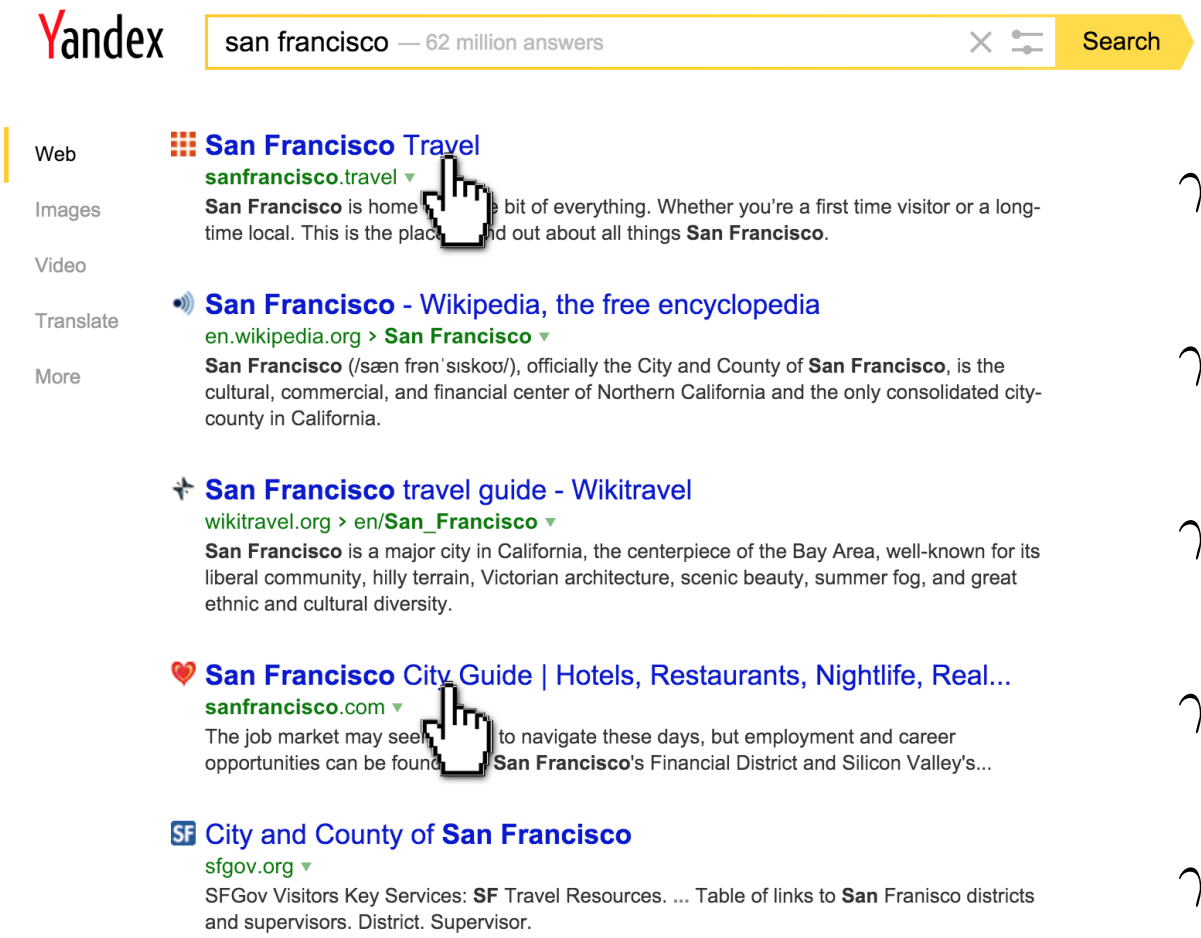
γ_5 , $P_{attr}(d_5 q)$

Position-based model: attractiveness

- Terminology
  - Attractiveness = a user wants to click on a document after examining (reading) its snippet
  - A_{qd} – binary random variable showing whether document d is attractive to a user, given query q
- Position-based model (PBM)
  - Attractiveness depends on a query-document pair

$$P(A_{qd} = 1) = \alpha_{qd}$$

Position-based model



γ_1 , α_{qd_1}

γ_2 , α_{qd_2}

γ_3 , α_{qd_3}

γ_4 , α_{qd_4}

γ_5 , α_{qd_5}

Position-based model: summary

$$P(E_{r_d} = 1) = \gamma_{r_d}$$

$$P(A_{qd} = 1) = \alpha_{qd}$$

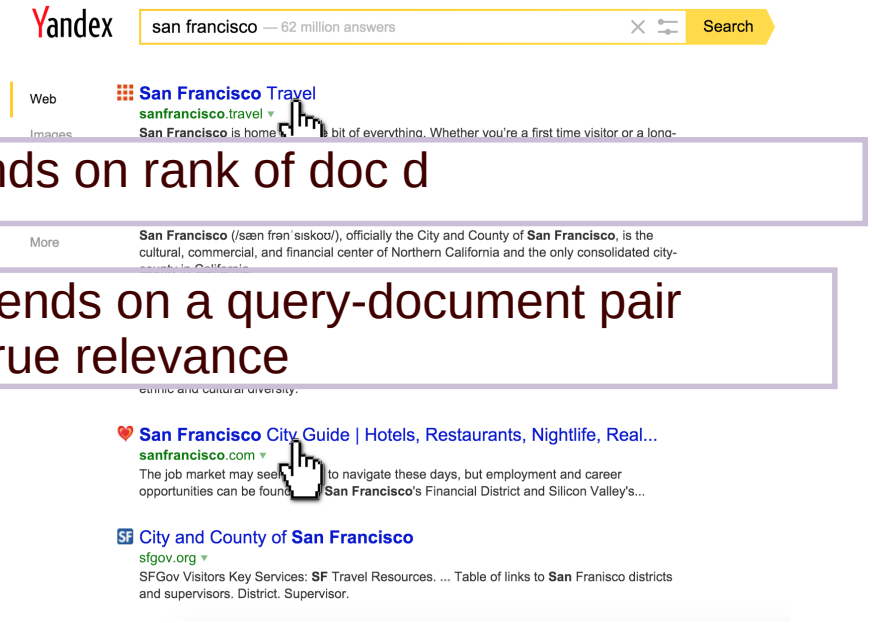
$$P(C_d = 1) = P(E_{r_d} = 1) \cdot P(A_{qd} = 1)$$

Examination depends on rank of doc d

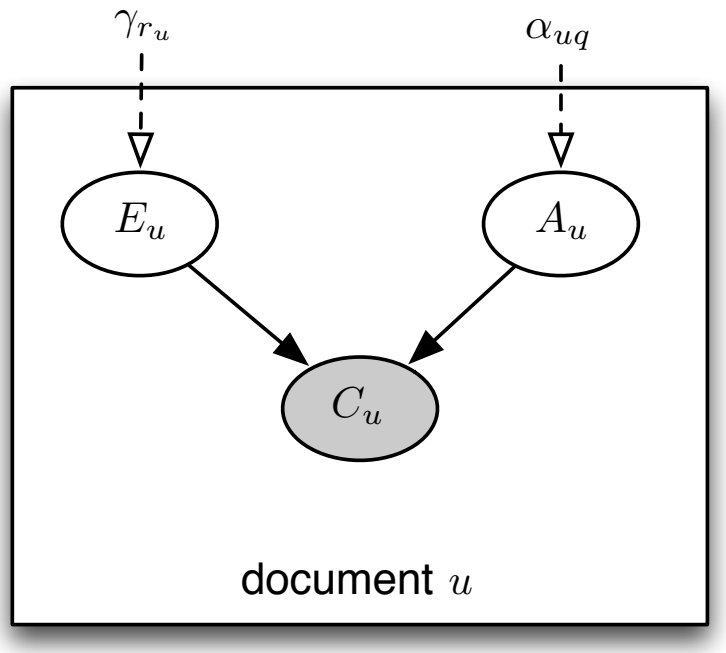
Attractiveness depends on a query-document pair
Attractiveness is true relevance

click is Biased

attractiveness is Unbiased



Position-based model: probabilistic graphical model

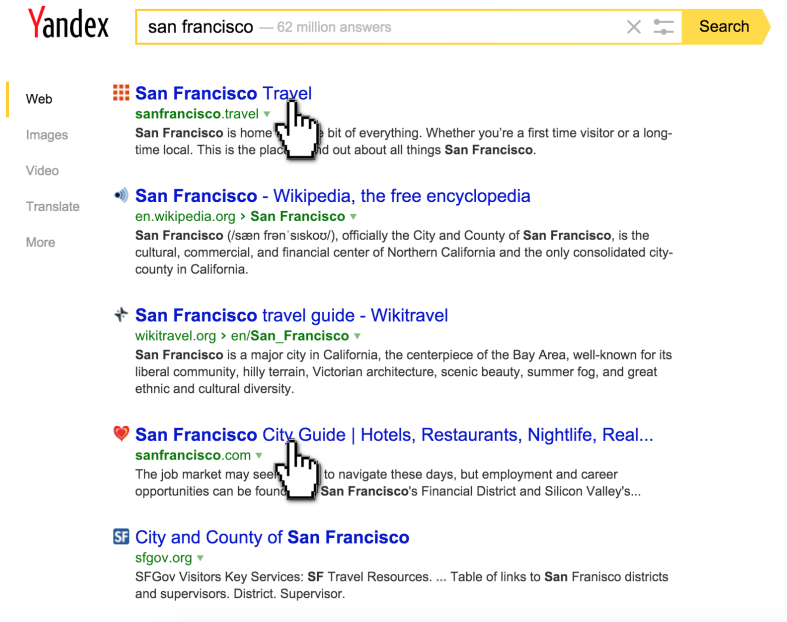


Position-based model

$$P(E_{r_d} = 1) = \gamma_{r_d}$$

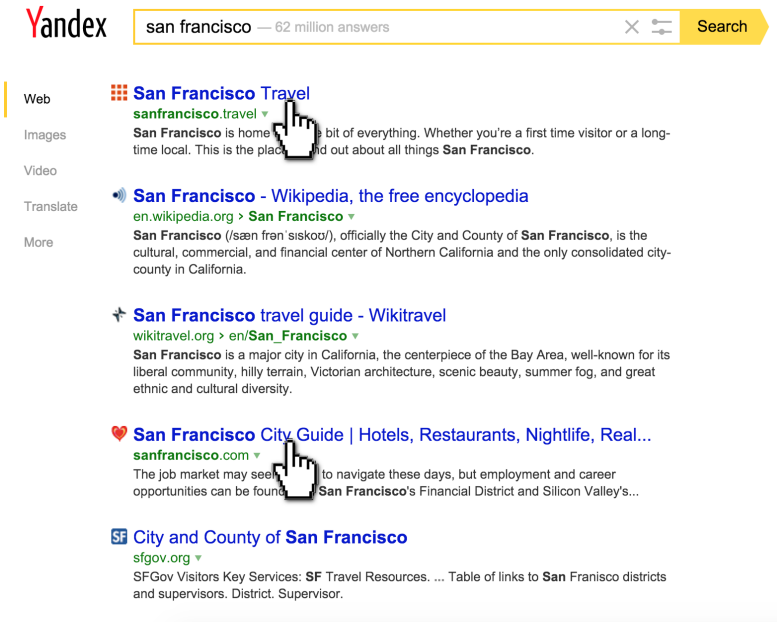
$$P(A_{qd} = 1) = \alpha_{qd}$$

$$P(C_d = 1) = P(E_{r_d} = 1) \cdot P(A_{qd} = 1)$$



Cascade model

1. Start from the first document
2. Examine documents one by one
3. If click, then stop
4. Otherwise, continue



Cascade model

we dont have para which depends on ranked position

$$E_r = 1 \text{ and } A_{d_r} = 1 \Leftrightarrow C_r = 1$$

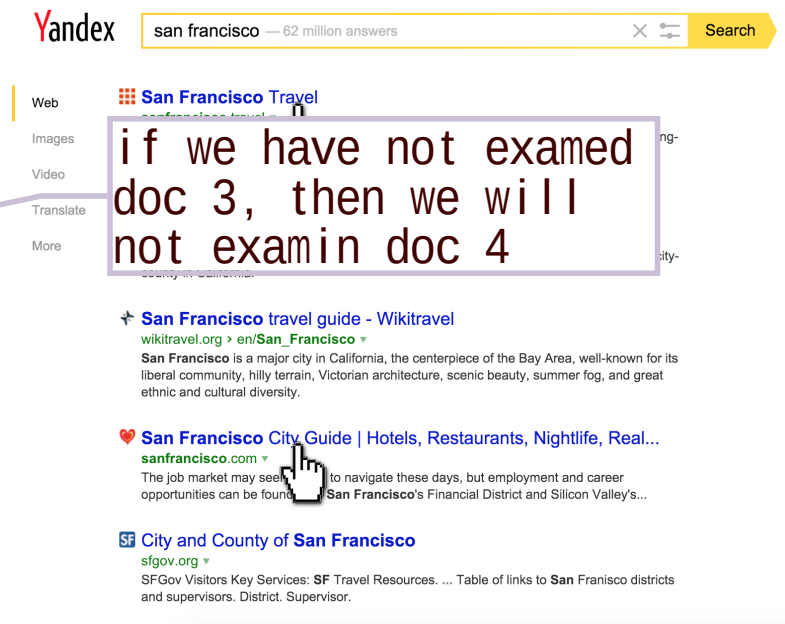
$$P(A_{d_r} = 1) = \alpha_{qd_r}$$

$$\underbrace{P(E_1 = 1)}_{\text{start from first}} = 1$$

$$\underbrace{P(E_r = 1 \mid E_{r-1} = 0)}_{\text{examine one by one}} = 0$$

$$\underbrace{P(E_r = 1 \mid C_{r-1} = 1)}_{\text{if click, then stop}} = 0$$

$$\underbrace{P(E_r = 1 \mid E_{r-1} = 1, C_{r-1} = 0)}_{\text{otherwise, continue}} = 1$$

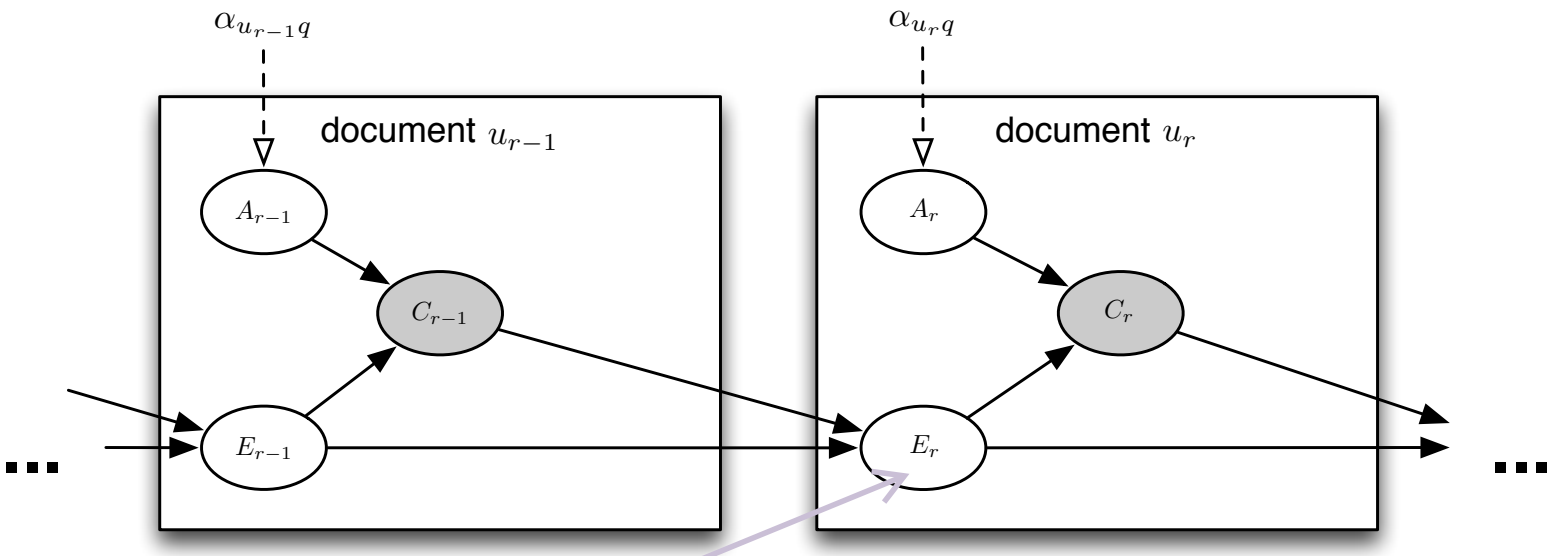


if we have not examed doc 3, then we will not examin doc 4

if we click doc 3, we don't exam doc 4

if we examined doc 3, and did not click doc 3, then we must examine doc 4

Cascade model: probabilistic graphical model



user will examine doc 3 only when:
the user has examined doc 2
and did not click doc 2

Basic click models summary

has bias

bias is removed
it is unbiased

- Position-based model (PBM)
  - \+ examination and attractiveness
  - – examination of a document at rank r does not depend on examinations and clicks above r
- Cascade model (CM)
  - \+ cascade dependency of examination at r on examinations and clicks above r

比较：
position based model: click happen in the previous time step has NO influence on whether the user will click the rest of snippets. they are indep. it only depends on the rank of snippet. PBM has 2 para: attractiveness, rank.

cascade: clicks happen above(排序靠前的snippet) influences whether user will click in the following snippet（排序靠后的snippets). they are dependent. PBM has 1 para: attractiveness. Cascade does not have rank para r

Outline

Parameter estimation

- Maximum likelihood estimation
- Expectation-maximization

Expectation maximization

1. Set parameters to some initial values
2. Repeat until convergence
   - E-step: derive the expectation of the likelihood function
   - M-step: maximize this expectation

Expectation maximization

$$Q(\theta_c) = \sum_{s \in \mathcal{S}} \mathbb{E}_{\mathrm{X}|\mathrm{C}^{(s)},\Psi} \left[\log P \left(\mathrm{X}, \mathrm{C}^{(s)} \mid \Psi \right) \right]$$

$$= \sum_{s \in \mathcal{S}} \mathbb{E}_{\mathrm{X}|\mathrm{C}^{(s)},\Psi} \left[\sum_{c_i \in s} \left(\mathcal{I} \left(X_{c_i}^{(s)} = 1, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \right) \log(\theta_c) + \mathcal{I} \left(X_{c_i}^{(s)} = 0, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \right) \log(1 - \theta_c) \right) + \mathcal{Z} \right]$$

$$= \sum_{s \in \mathcal{S}} \sum_{c_i \in s} \left(P \left(X_{c_i}^{(s)} = 1, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right) \log(\theta_c) + P \left(X_{c_i}^{(s)} = 0, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right) \log(1 - \theta_c) \right) + \mathcal{Z}$$

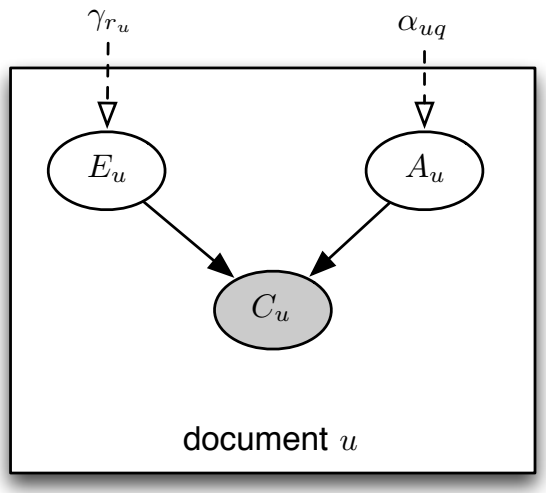
$$ESS(x) = \sum_{s \in \mathcal{S}} \sum_{c_i \in s} P \left(X_{c_i}^{(s)} = x, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)$$

$$\frac{\partial Q(\theta_c)}{\partial \theta_c} = \sum_{s \in \mathcal{S}} \sum_{c_i \in s} \left(\frac{P \left(X_{c_i}^{(s)} = 1, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)}{\theta_c} - \frac{P \left(X_{c_i}^{(s)} = 0, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)}{1 - \theta_c} \right) = 0$$

$$\theta_c^{(t+1)} = \frac{\sum_{s \in \mathcal{S}} \sum_{c_i \in s} P \left(X_{c_i}^{(s)} = 1, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)}{\sum_{s \in \mathcal{S}} \sum_{c_i \in s} \sum_{x=0}^{x=1} P \left(X_{c_i}^{(s)} = x, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)}$$

$$= \frac{\sum_{s \in \mathcal{S}} \sum_{c_i \in s} P \left(X_{c_i}^{(s)} = 1, \mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)}{\sum_{s \in \mathcal{S}} \sum_{c_i \in s} P \left(\mathcal{P}(X_{c_i}^{(s)}) = \mathrm{p} \mid \mathrm{C}^{(s)}, \Psi \right)} = \frac{ESS^{(t)}(1)}{ESS^{(t)}(1) + ESS^{(t)}(0)}$$

EM for Position-Based Model



$$P(A_d = 1) = \alpha_{qd}$$
$$P(E_r = 1) = \gamma_r$$

EM update rules for PBM: attractiveness

$$\alpha_{qd}^{(t+1)} = \frac{1}{|\mathcal{S}_{qd}|} \sum_{s \in \mathcal{S}_{qd}} \left(c_d^{(s)} + (1 - c_d^{(s)}) \frac{(1 - \gamma_r^{(t)}) \alpha_{qd}^{(t)}}{1 - \gamma_r^{(t)} \alpha_{qd}^{(t)}} \right)$$

- t – iteration
- $\mathcal{S}_{qd}$ – search sessions initiated by query q and containing document u
- $c_d^{(s)}$ – observed click on document u in search session s

EM update rules for PBM: examination

$$\gamma_r^{(t+1)} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left(c_d^{(s)} + (1 - c_d^{(s)}) \frac{\gamma_r^{(t)} (1 - \alpha_{qd}^{(t)})}{1 - \gamma_r^{(t)} \alpha_{qd}^{(t)}} \right)$$

Outline

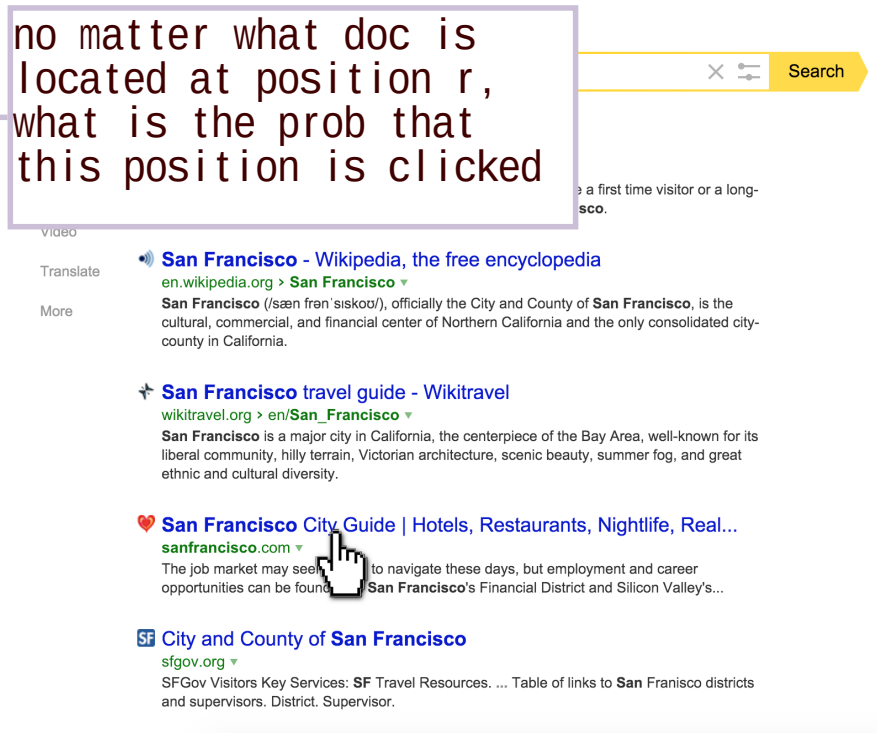
Click probabilities

- Full probability – probability that a user clicks on a document at rank r

$$P(C_r = 1)$$

- Conditional probability – probability that a user clicks on a document at rank r given previous clicks

$$P(C_r = 1 \mid C_1, \ldots, C_{r-1})$$



Applications of click models

RBP and ERR belongs to this.
they are based on user based models

Click model's output	Application
Full click probabilities	Model-based metrics
Conditional click probabilities	User simulation
Parameter values	Ranking

Model-based metrics

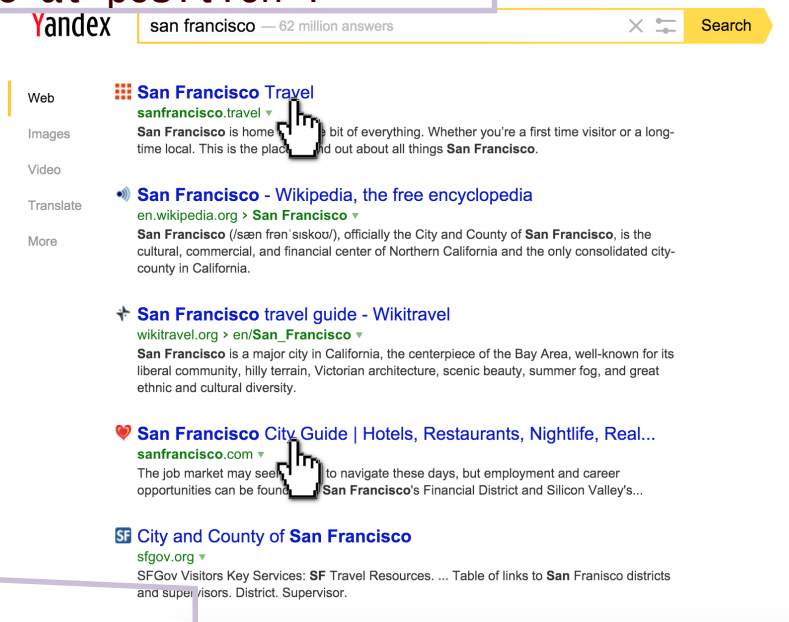
- Utility-based metrics

$$uMetric = \sum_{r=1}^{n} P(C_r = 1) \cdot U_r$$

- Effort-based metrics

$$eMetric = \sum_{r=1}^{n} P(S_r = 1) \cdot F_r$$

U: utility=happiness of user, after click the doc at position r



efforts that user make to get satisfication from the doc which rank at r positon

Expected reciprocal rank

satisfication =1 or 0
=1 when user is satisfied with the doc, after click it

r: rank of a doc

$$RR = \frac{1}{r}, \text{ where } S_r = 1$$

$$ERR = \sum_r \frac{1}{r} \cdot P(S_r = 1)$$

sum over all doc that user is satisfied after click them

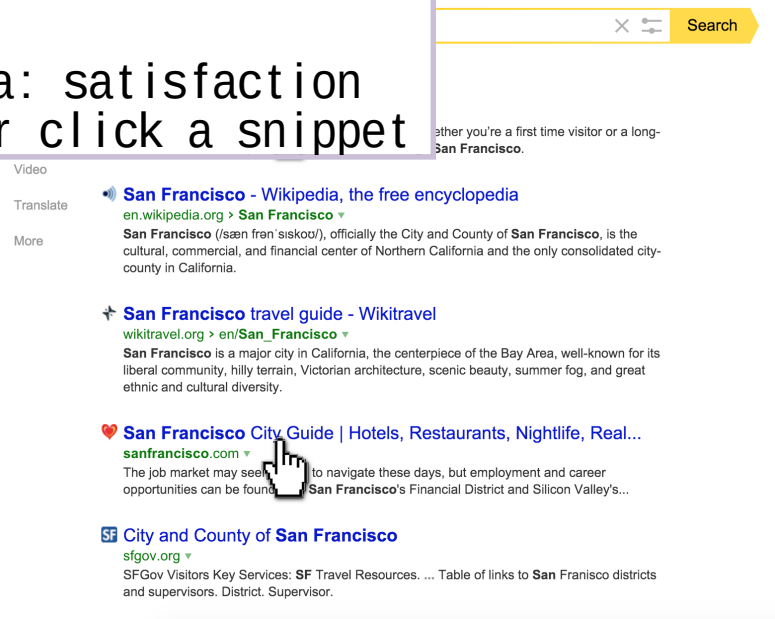
Dynamic Bayesian network model (DBN)

$$P(A_r = 1) = \alpha_{qd_r}$$
$$P(E_1 = 1) = 1$$
$$P(E_r = 1 \mid S_{r-1} = 1) = 0$$
$$P(E_r = 1 \mid S_{r-1} = 0) = \gamma$$
$$P(S_r = 1 \mid C_r = 0) = 0$$
$$P(S_r = 1 \mid C_r = 1) = \sigma_{qd_r}$$

$$P(S_r = 1) = ?$$

DBN has 2 satisfaction para:
alpha: satisfaction after reading the snippet

sigma: satisfaction after click a snippet



DBN: Satisfaction

$$
\begin{aligned}
P(S_r = 1) &= P(S_r = 1 \mid C_r = 1) \cdot P(C_r = 1) \\
&= \sigma_{qd_r} \cdot P(C_r = 1) \\
&= \sigma_{qd_r} \cdot \alpha_{qd_r} \cdot P(E_r = 1) \\
&= \sigma_{qd_r} \cdot \alpha_{qd_r} \cdot \prod_{i=1}^{r-1} \left(\gamma \cdot (1 - \sigma_{qd_i} \cdot \alpha_{qd_i})\right) \\
&= R_{qd_r} \cdot \prod_{i=1}^{r-1} \left(\gamma \cdot (1 - R_{qd_i})\right)
\end{aligned}
$$

Expected reciprocal rank

$$ERR = \sum_r \frac{1}{r} \cdot P(S_r = 1)$$

$$= \sum_r \frac{1}{r} \cdot R_{qd_r} \cdot \prod_{i=1}^{r-1} \left(\gamma \cdot (1 - R_{qd_i})\right)$$

Summary

- Interactions
  - Examples of interactions
  - Applications: evaluate and improve IR
- Interaction models
  - Basic click models: PBM, CM
  - Applications: ERR

Materials

- Aleksandr Chuklin, Ilya Markov, Maarten and de Rijke
  Click Models for Web Search
  Morgan & Claypool, 2015