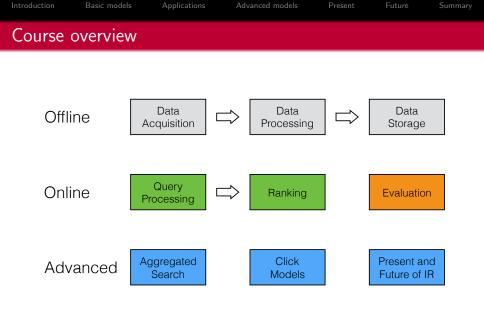
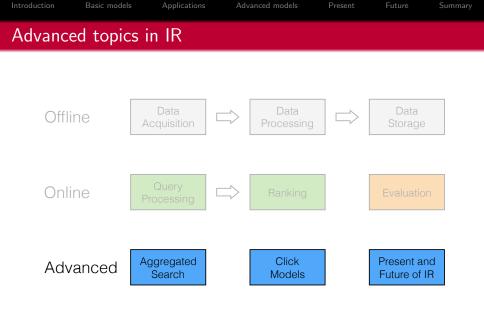
Introduction	Basic models	Applications	Advanced models	Present	Future	Summary

# Information Retrieval Click Models

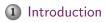
### Ilya Markov i.markov@uva.nl

University of Amsterdam



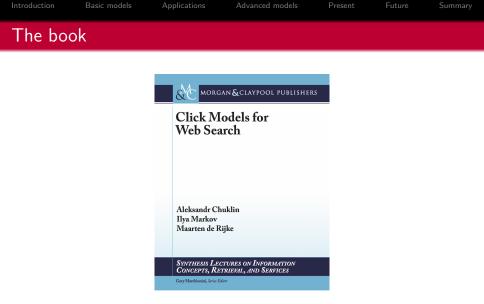


Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						



- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research





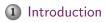
http://clickmodels.weebly.com/the-book.html

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Tutorials	5					

- SIGIR 2015, Santiago, Chile
- AINL-ISMW FRUCT 2015, St. Petersburg, Russia
- WSDM 2016, San Francisco, USA
- RuSSIR 2016, Saratov, Russia

http://clickmodels.weebly.com/tutorials.html

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						



- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research

## 7 Summary

Ilya Markov

Introduction

Basic models

Applications

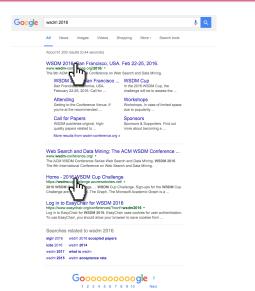
Advanced models

Present

Future

Summary

# Why click models?



Introduction Basic models Applications Advanced models Present Future Summary
Why click models?

Scientific modelling is a scientific activity, the aim of which is to make a particular part or feature of the world easier to understand, define, quantify, visualize, or simulate by referencing it to existing and usually commonly accepted knowledge.

Wikipedia, Scientific modelling

Introduction Basic models Applications Advanced models Present Future Summary
Why click models?

Click models make user clicks in web search easier to understand, define, quantify, visualize, or simulate using (mostly) probabilistic graphical models.

$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Introduction		Basic mo	dels	Арр			Advance	ed mode	ls	Preser		Future		Summary	
0       36       0       174       0       1625       1627       1623       1626       1624       1619       1621       1620       1618         0       50       0       227       0       2091       2087       2087       2083       2088       2092       2092       2095       2086         0       515       0       174       0       1625       1627       1623       1626       1624       1622       1619       1621       1620       1618         0       524       0       1974       0       1752       1627       1626       1624       1622       1619       1621       1620       1618         0       527       0       1744       0       1625       1627       1623       2091       1753       1753       1756       1756       1756         0       529       C       1626       -       -       -       -       -       -       -       -       -       -       -       -       1623       181       104       29       21       89       85         1       0       0       9       1807       11807       11805 <td< th=""><th>Click I</th><th>og</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></td<>	Click I	og														
0       36       0       174       0       1625       1627       1623       1626       1624       1619       1621       1620       1618         0       50       0       227       0       2091       2087       2087       2083       2088       2092       2092       2095       2086         0       515       0       174       0       1625       1627       1623       1626       1624       1622       1619       1621       1620       1618         0       524       0       1974       0       1752       1627       1626       1624       1622       1619       1621       1620       1618         0       527       0       1744       0       1625       1627       1623       2091       1753       1753       1756       1756       1756         0       529       C       1626       -       -       -       -       -       -       -       -       -       -       -       -       1623       181       104       29       21       89       85         1       0       0       9       1807       11807       11805 <td< th=""><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></td<>																
1       0       9       0       13       70       66       94       50       104       29       21       89       85         1       20       C       104       123       C       21       104       1107       11805       11812       11813       11809       11809       11810       11811       11808       11810         1       291       Q       1324       0       11807       11805       11812       11813       11809       11806       11811       11808       11810         1       301       C       11813       11807       11813       11804       11809       11805       11811       11808       11810         1       8605       C       11818       11810       11810       11813       11804       11809       11805       11811       11808       11810         1       8605       C       11810       11810       11810       11811       11808       11810         1       8737       C       11810       11811       11808       11811       11805       1181       11810       11810       11810       11810       11810       11810       11810       11810 <th>0 0 0 0 0</th> <th>36 50 515 524 527 528</th> <th>Q Q Q C C</th> <th>174 227 174 1974 17562 1627</th> <th>0 0 0</th> <th>1625 2094 1625</th> <th>1627 2091 1627</th> <th>1623 2087 1623</th> <th>1626 2089 1626</th> <th>1624 2093 1624</th> <th>1622 2088 1622</th> <th>1619 2090 1619</th> <th>1621 2092 1621</th> <th>1620 2095 1620</th> <th>1618 2086 1618</th> <th></th>	0 0 0 0 0	36 50 515 524 527 528	Q Q Q C C	174 227 174 1974 17562 1627	0 0 0	1625 2094 1625	1627 2091 1627	1623 2087 1623	1626 2089 1626	1624 2093 1624	1622 2088 1622	1619 2090 1619	1621 2092 1621	1620 2095 1620	1618 2086 1618	
1       291       0       1324       0       11807       11805       11812       11813       11804       11809       11806       11811       11808       11810         1       301       C       11803       11810       11800       11810       11810       11810	1 1	0 20	Q C	9 104	0	13	70	66	94	50	104	29	21	89	85	
2 0 0 7 0 77 93 55 86 64 67 76 98 18 54 2 11 C 18 2 1122 0 4088 0 35554 35561 35562 35556 35557 35567 35550 35566 35568 35553 2 1127 C 35561 2 1645 Q 5663 0 36505 36514 36508 36509 50480 36510 36507 50482 50483 50481 2 1646 C 35695	1 1 1 1	291 301 8605 8737	Q C C C	1324 11813 11808 11810	0	11807	11805	11812	11813	11804	11809	11806	11811	11808	11810	
2 11 C 18 2 1122 Q 4088 0 35554 35561 35562 35556 35557 35567 35550 35566 35568 35553 2 1127 C 35561 2 1645 Q 5663 0 36505 36514 36508 36509 50480 36510 36507 50482 50483 50481 2 1646 C 35695	2	0			0	77	93	55	86	64	67	76	98	18	54	
2 1122 C 35561 5356 5356 5357 5357 5357 5357 5358 5358 5358 5358	2				0	25554	25561	25562	25556	25557	25567	25550	25566	25560	25552	
2 1645 Q 5863 Ø 36505 36514 36508 36509 50480 36510 36507 50482 50483 50481 2 1646 C 36505	2				U	33354	22201	33562	33350	33557	33567	33320	33300	33268	30003	
	2	1645	Q	5863	0	36505	36514	36508	36509	50480	36510	36507	50482	50483	50481	

Yandex Relevance Prediction Challenge http://imat-relpred.yandex.ru/en

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

## 1 Introduction

## 2 Basic click models

- Random click model
- Position-based model
- Cascade model
- Click probabilities
- Evaluation
- Parameter estimation

## 3 Applications

## 4 Advanced models

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

## 2 Basic click models

### Random click model

- Position-based model
- Cascade model
- Click probabilities
- Evaluation
- Parameter estimation

	ro			

Basic models

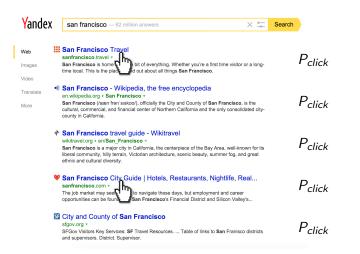
Applications

Advanced models

Present Future

Summary

# Random click model



- Terminology
  - $C_u$  binary random variable denoting a click on document u
- Random click model (RCM)
  - Any document can be clicked with the same (fixed) probability

$$P(C_u = 1) = const = \rho$$

Web

Images

Video

More

Advanced models

Summary

# Random click model

Yandex san francisco — 62 million	n answ
-----------------------------------	--------

San Francisco Travel

time local. This is the place

sanfrancisco.travel v

🗙 🚍 🛛 Search

 $P(C_{u_1}=1)=\rho$ 

 $P(C_{\mu_2} = 1) = \rho$ 

 $P(C_{\mu_2} = 1) = \rho$ 

 $P(C_{\mu} = 1) = \rho$ 

 $P(C_{\mu \epsilon} = 1) = \rho$ 

#### San Francisco - Wikipedia, the free encyclopedia en.wikipedia.org > San Francisco \*

San Francisco (/sæn fren'siskoo/), officially the City and County of San Francisco, is the cultural, commercial, and financial center of Northern California and the only consolidated citycounty in California.

nd out about all things San Francisco

bit of everything. Whether you're a first time visitor or a long-

### \* San Francisco travel guide - Wikitravel

#### wikitravel.org > en/San\_Francisco \*

San Francisco is a major city in California, the centerpiece of the Bay Area, well-known for its liberal community, hilly terrain, Victorian architecture, scenic beauty, summer fog, and great ethnic and cultural diversity.

### San Francisco City Guide | Hotels, Restaurants, Nightlife, Real...

The job market may seen to navigate these days, but employment and career opportunities can be found. San Francisco's Financial District and Silicon Valley's...

### City and County of San Francisco

#### sfgov.org •

SFGov Visitors Key Services: SF Travel Resources. ... Table of links to San Franisco districts and supervisors. District. Supervisor.

$$\rho = \frac{\# \text{ clicks}}{\# \text{ shown docs}}$$

Random click model (global CTR):

$$P(C_u=1)=\rho$$

Rank-based CTR:

$$P(C_{u_r}=1)=\rho_r$$

Query-document CTR:

$$P(C_u=1)=\rho_{uq}$$

<b>Y</b> ande:	san francisco — 62 million answers X 🛻	Search
Web	san Francisco Travel	
Images Video	San Francisco is home to be of everything. Whether you're a first time visitor or a long- time local. This is the place and out about all things San Francisco.	
Translate	<ul> <li>San Francisco - Wikipedia, the free encyclopedia en.wikipedia.org &gt; San Francisco *</li> </ul>	
More	San Francisco (Isan frantisskost), officially the City and County of San Francisco, is the cultural, commercial, and financial center of Northern Celifornia and the only consolidated oly- county in Celifornia.	
	San Francisco travel guido - Wikitravel wk/travel.cgp - on/3an_Francisco * Ban Francisco a major dy in Calornia, the centerpiece of the Bay Area, well-known for its found commany, hitly larger, Victorian achitecture, science basity, science fog, and great entries card calored eventy.	
	San Francisco City Guide   Hotels, Restaurants, Nightlife, Real sanfrancisco.com * The job market way seen Diso navigate these days, but employment and career opportunities can be found than Francisco's Financial District and Silcon Valley's	
	City and County of San Francisco slov up * SFGor Vialow Key Santoas: SF Travel Resources Table of Inks to San Frankoo districts and supervisors. Dathol. Supervisor.	

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

## 2 Basic click models

- Random click model
- Position-based model
- Cascade model
- Click probabilities
- Evaluation
- Parameter estimation

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Position	n-based mo	odel				
Yande	san francisco — 62 mill	ion answers	X 🛫 Sea	arch		
Web Images	San Francisco Travel sanfrancisco.travel • San Francisco is home time local. This is the place	bit of everything. Whether you're	a first time visitor or a long- sco.	${\sf P}_{{\it read}}(1)$ , i	$P_{click}(u_1q)$	)
Video Translate More			of San Francisco, is the	P <sub>read</sub> (2), F	$P_{click}(u_2q)$	
				$P_{read}(3)$ , $P_{read}(3)$ , $P_{read}(3)$ , $P_{read}(3)$	P <sub>click</sub> (u <sub>3</sub> q	)
		uide   Hotels, Restaurant to navigate these days, but empl San Francisco's Financial Distric	pyment and career	$P_{read}(4)$ , $I$	P <sub>click</sub> (u <sub>4</sub> q	)
	SI City and County of San	Francisco			<b>D</b> (	<b>、</b>

### sfgov.org v

SFGov Visitors Key Services: SF Travel Resources. ... Table of links to San Franisco districts and supervisors. District. Supervisor.  $P_{\mathit{read}}(5)$  ,  $P_{\mathit{click}}(u_5q)$ 

# 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 •

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

 $\times =$ 

Search

Yandex san francisco — 62 million answers



 $\gamma_1$ ,  $P_{click}(u_1q)$ 

 $\gamma_2$ ,  $P_{click}(u_2q)$ 

#### \* San Francisco travel guide - Wikitravel wikitravel.org > en/San Francisco v

San Francisco is a major city in California, the centerpiece of the Bay Area, well-known for its liberal community, hilly terrain, Victorian architecture, scenic beauty, summer fog, and great ethnic and cultural diversity.

#### San Francisco City Guide | Hotels, Restaurants, Nightlife, Real... sanfrancisco.com \*

to navigate these days, but employment and career The job market may see San Francisco's Financial District and Silicon Valley's... opportunities can be found

### SI City and County of San Francisco

#### sfaoy.org v

SFGov Visitors Key Services: SF Travel Resources. ... Table of links to San Franisco districts and supervisors. District. Supervisor.

 $\gamma_3$ ,  $P_{click}(u_3q)$ 

 $\gamma_4$ ,  $P_{click}(u_4q)$ 

 $\gamma_5$ ,  $P_{click}(u_5q)$ 

Web

# Position-based model: attractiveness

## Terminology

- Attractiveness = a user wants to click on a document after examining (reading) its snippet
- $A_{\mu}$  binary random variable showing whether document u is attractive to a user, given query q

Position-based model (PBM)

Attractiveness depends on a guery-document pair

$$P(A_{uq}=1)=lpha_{uq}$$

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Position	-based mo	del				
Yande	san francisco — 62 n	illion answers	X 🚍 Search			
Web Images Video	San Francisco Trave sanfrancisco.travel v San Francisco is home time local. This is the place	) h bit of everything. Whether you do out about all things San Fran		$\gamma_1$ , $lpha_{\mathit{u}_1 \mathit{q}}$		
Translate	en.wikipedia.org > San Fra San Francisco (/sæn fran'se	pedia, the free encyclope ncisco v kow/), officially the City and County ncial center of Northern California	/ of San Francisco, is the	$\gamma_2$ , $lpha_{\it u_2 \it q}$		
				$\gamma_{ m 3}$ , $lpha_{{\it u}_{ m 3}}{\it q}$		
	San Francisco City sanfrancisco.com • The job market may seen opportunities can be found	Guide   Hotels, Restauran to navigate these days, but em San Francisco's Financial Distr	ployment and career	$\gamma_{ m 4}$ , $lpha_{\it u_{ m 4}\it q}$		
	SE City and County of Se sfgov.org * SFGov Visitors Key Services: and supervisors. District. Sup	SF Travel Resources Table of	links to San Franisco districts	$\gamma_5$ , $lpha_{\mathit{U}_5 \mathit{q}}$		

Basic models

Applications

Advanced models

Present

: Fut

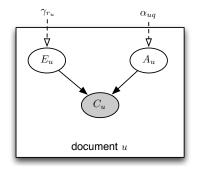
Summary

# Position-based model: summary

$$\begin{split} P(E_{r_{u}} = 1) &= \gamma_{r_{u}} \\ P(A_{u} = 1) &= \alpha_{uq} \\ P(C_{u} = 1) &= P(E_{r_{u}} = 1) \cdot P(A_{u} = 1) \end{split}$$

 Introduction
 Basic models
 Applications
 Advanced models
 Present
 Future
 Summary

 Position-based model:
 probabilistic graphical model
 Model
 Present
 Future
 Summary



Applications

Advanced models

Present

nt Future

Summary

## Position-based model: exercises

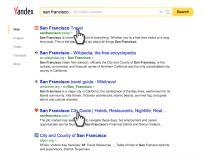
$$P(E_{r_u} = 1) = \gamma_{r_u}$$
  
 $P(A_u = 1) = \alpha_{uq}$   
 $P(C_u = 1) = P(E_{r_u} = 1) \cdot P(A_u = 1)$ 

$$E_{r_u} = 0 \Rightarrow C_u = 0$$

$$A_u = 0 \Rightarrow C_u = 0$$

$$E_{r_u} = 1 \Rightarrow (C_u = 1 \iff A_u = 1)$$

$$A_u = 1 \Rightarrow (C_u = 1 \iff E_{r_u} = 1)$$



Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

## 2 Basic click models

- Random click model
- Position-based model

### Cascade model

- Click probabilities
- Evaluation
- Parameter estimation

Advanced models

Present

e Su

# Position-based model

$$\begin{aligned} & P(E_{r_{u}} = 1) = \gamma_{r_{u}} \\ & P(A_{u} = 1) = \alpha_{uq} \\ & P(C_{u} = 1) = P(E_{r_{u}} = 1) \cdot P(A_{u} = 1) \end{aligned}$$

Yandex	san francisco — 62 million answers	$\times \doteqdot$	Search
Web	San Francisco Travel sanfrancisco.travel *		
Images	San Francisco is home bit of everything. Whether you're a first time visitor time local. This is the place of out about all things San Francisco.	or a long-	
Video			
Translate	San Francisco - Wikipedia, the free encyclopedia en.wikipedia.org > San Francisco +		
More	San Francisco (saan fran siskoor), officially the City and County of San Francisco, cultural, commercial, and financial center of Northern California and the only consoli county in California.		
	San Francisco travel guide - Wikitravel wkitravel.org > cn/San_Francisco + San Francisco is a major toty in California, the centerpiece of the Bay Area, well-ien beard community, high senait, Netsina authideure, scenic beauty, summar log, ar		
	ethnic and cultural diversity.		
	San Francisco City Guide   Hotels, Restaurants, Nightlife, R samfrancisco.com - The job market may see the provided of the pro	r	
	City and County of San Francisco sfgov.org + SFGov Values Key Services: SF Toront Resources Table of Interto San Francisc	n districts	
	and supervisors. District. Supervisor.		

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

Yande>	san francisco — 82 million answers	× =	Search
Web	San Francisco Travel		
Images	San Francisco // avei + San Francisco is home time local. This is the place of out about all things San R	you're a first time visitor or a long- Francisco.	
Video Translate	San Francisco - Wikipedia, the free encycl	lopedia	
More	en.wikipedia.org > San Francisco + San Francisco (san fran'ssikool), officially the City and Ci cultural, commercial, and financial center of Northern Califor county in California.		
	San Francisco travel guide - Wikitravel wikitravel.org > en/San_Francisco * San Francisco is a major oty in California, the centerpiece liberal community, hilly termin, Victorian architecture, acento ethnic and cultural diversity.		
	San Francisco City Guide   Hotels, Restau sanfrancisco.com * The job market may see to navigate these days, bu opportunities can be found San Francisco's Financial	t employment and career	
	City and County of San Francisco signv.org v SFGov Visitors Key Services: SF Travel Resources Table	e of links to San Franisco districts	

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Cascade	e model					

$$E_{r} = 1 \text{ and } A_{u_{r}} = 1 \Leftrightarrow C_{r} = 1$$

$$P(A_{u_{r}} = 1) = \alpha_{u_{r}q}$$

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

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

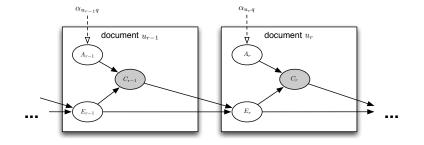
$$\underbrace{P(E_{r} = 1 \mid C_{r-1} = 1)}_{\text{if click, then stop}} = 0$$

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

× 🚎 Search

bit of everything. Whether you're a first time visitor or a long-Id out about all things San Francisco.





# Basic click models summary

## CTR models

- + count clicks (simple and fast)
- do not distinguish examination and attractiveness
- - + examination and attractiveness
  - examination of a document at rank r does not depend on examinations and clicks above r
- Cascade model (CM) Dynamic Bayesian network
  - + cascade dependency of examination at ron examinations and clicks above r
  - only one click is allowed

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

### 2 Basic click models

- Random click model
- Position-based model
- Cascade model

## Click probabilities

- Evaluation
- Parameter estimation

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Click pr	obabilities					

• 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 rgiven previous clicks

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

Yande	San francisco – 62 million anewers X	Search
Web	San Francisco Travel	
Images	San Francisco is home bit of everything. Whether you're a first time visitor or a long- time local. This is the place of our about all things San Francisco.	
Video Translate	San Francisco - Wikipedia, the free encyclopedia en wkipedia.crg > San Francisco +	
More	en.wwnpiera.org 9 sum Prancisco * San Francisco (seen fren'ssicos), officially the City and County of San Francisco, is the cultural, commercial, and financial center of Northern California and the only consolidated city- county in California.	
	★ San Francisco travel guide - Wikitravel wikitravel.org > enSan_Francisco * San Francisco is amport of in callomia, the centerpiece of the Bay Area, well-incen for its Iberal community, INIy terrait, Victorian architecture, scenic beauty, summer for, and great ethnic and output of interpiece.	
	San Francisco City Guide   Hotels, Restaurants, Nightlife, Real     sanfrancisco com *     The job market may soon     for navigate these days, but employment and career     opportunities can be found     San Francisco's Financial District and Sticon Valley's	
	City and County of San Francisco slopv.org * SFGov Valiors Kay Services: SF Travel Resources Table of links to San Franisco districts and supervisor. Dahrd. Supervisor.	

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Click pr	obabilities					

• Full probability

$$P(C_{r+1} = 1) =$$
  

$$\alpha_{u_{r+1}q}\epsilon_r \cdot \begin{pmatrix} P(E_{r+1} = 1 \mid E_r = 1, C_r = 1) \cdot P(C_r = 1 \mid E_r = 1) \\ P(E_{r+1} = 1 \mid E_r = 1, C_r = 0) \cdot P(C_r = 0 \mid E_r = 1) \end{pmatrix}$$

Conditional probability

$$P(C_{r+1} = 1 | C_1, ..., C_r)$$

$$= \alpha_{u_{r+1}q} \cdot \begin{pmatrix} P(E_{r+1} = 1 | E_r = 1, C_r = 1) \cdot c_r^{(s)} \\ + P(E_{r+1} = 1 | E_r = 1, C_r = 0) \cdot \frac{\epsilon_r (1 - \alpha_{u_rq})}{1 - \alpha_{u_rq}\epsilon_r} \cdot (1 - c_r^{(s)}) \end{pmatrix}$$

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

### 2 Basic click models

- Random click model
- Position-based model
- Cascade model
- Click probabilities

### Evaluation

Parameter estimation

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Evaluat	ion					

Click model's output	Evaluation
Full click probabilities	Perplexity
Conditional click probabilities	Log-likelihood



Perplexity measures how well a click model estimates full click probabilities (i.e., when clicks are not observed).

$$p_r(M) = 2^{-\frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}} \left(\log_2 \frac{p_M(C_r^{(s)} = c_r^{(s)})}{p_M(C_r^{(s)} = c_r^{(s)})}\right)}$$
$$p_r(M) \in [1..2]$$



Likelihood measures how well a click model estimates conditional click probabilities given observed clicks.

$$\mathcal{LL}(M) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \log P_M \left( C_1 = c_1^{(s)}, \dots, C_n = c_n^{(s)} \right)$$
$$= \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \sum_{r=1}^n \log \underbrace{P_M \left( C_r = c_r^{(s)} \mid \mathbf{C}_{< r} = \mathbf{c}_{< r}^{(s)} \right)}_{\text{conditional click probability}}$$

$$\mathcal{LL}(M) \in [-\infty..0]$$

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

#### 2 Basic click models

- Random click model
- Position-based model
- Cascade model
- Click probabilities
- Evaluation

#### Parameter estimation

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Parame	ter estima	tion				

- Maximum likelihood estimation
- Expectation-maximization

Introduction Basic models Applications Advanced models Present Future Summary
MLE for random click model

$$P(C_u=1)=\rho$$

$$\mathcal{L} = \prod_{s \in S} \prod_{u \in s} \rho^{c_u^{(s)}} (1 - \rho)^{1 - c_u^{(s)}}$$
likelihood of Bernoulli random variable

$$\mathcal{LL} = \sum_{s \in \mathcal{S}} \sum_{u \in s} \left( c_u^{(s)} \log(
ho) + (1 - c_u^{(s)}) \log(1 - 
ho) 
ight)$$

$$\rho = \frac{\sum_{s \in \mathcal{S}} \sum_{u \in s} c_u^{(s)}}{\sum_{s \in \mathcal{S}} |s|} = \frac{\# \text{ clicks}}{\# \text{ shown docs}}$$



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

Introduction Basic models Applications Advanced models Present Future Summary

#### Expectation maximization

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

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

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

$$\begin{split} \theta_{c}^{(t+1)} &= \frac{\sum_{s \in S} \sum_{c_{i} \in s} P\left(X_{c_{i}}^{(s)} = 1, \mathcal{P}(X_{c_{i}}^{(s)}) = \mathbf{p} \mid \mathbf{C}^{(s)}, \mathbf{\Psi}\right)}{\sum_{s \in S} \sum_{c_{i} \in s} \sum_{x=0}^{x=1} P\left(X_{c_{i}}^{(s)} = x, \mathcal{P}(X_{c_{i}}^{(s)}) = \mathbf{p} \mid \mathbf{C}^{(s)}, \mathbf{\Psi}\right)} \\ &= \frac{\sum_{s \in S} \sum_{c_{i} \in s} P\left(X_{c_{i}}^{(s)} = 1, \mathcal{P}(X_{c_{i}}^{(s)}) = \mathbf{p} \mid \mathbf{C}^{(s)}, \mathbf{\Psi}\right)}{\sum_{s \in S} \sum_{c_{i} \in s} P\left(\mathcal{P}(X_{c_{i}}^{(s)}) = \mathbf{p} \mid \mathbf{C}^{(s)}, \mathbf{\Psi}\right)} = \frac{ESS^{(t)}(1)}{ESS^{(t)}(1) + ESS^{(t)}(0)} \end{split}$$



- Basic click models
  - CTR models
  - Position-based model
  - Cascade model
- Click probabilities
  - Full click probabilities
  - Conditional click probabilities
- Evaluation
  - Perplexity
  - Log-likelihood
- Parameter estimation
  - Maximum likelihood estimation
  - Expectation-maximization



#### General:

Understanding of user behavior

Specific:

- Conditional click probabilities
- Full click probabilities
- Attractiveness and satisfactoriness for query-document pairs

### Applications

Click model's output	Application
Understanding of user behavior	User interaction analysis
Conditional click probabilities	User simulation
Full click probabilities	Model-based metrics
Parameter values	Ranking

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						



#### 2 Basic click models



- 4 Advanced models
- 5 Current developments
- 6 Future research

#### 7 Summary

Ilya Markov



- Random click model (global CTR):  $\rho = 0.122$
- Rank-based CTR:

 $\rho_1 = 0.429, \rho_2 = 0.190, \rho_3 = 0.136, \dots, \rho_{10} = 0.048$ 

- Position-based model:  $\gamma_1 = 0.998, \gamma_2 = 0.939, \gamma_3 = 0.759, \dots, \gamma_{10} = 0.260$
- Dynamic Bayesian network model:  $\gamma = 0.9997$

Click models are trained on the first 10K sessions of the WSCD 2012 dataset.

**Algorithm** Simulating user clicks Input: click model M, query session sOutput: vector of simulated clicks  $(c_1, \ldots, c_n)$ 

1: for 
$$r \leftarrow 1$$
 to  $|s|$  do  
2:  $P_r \leftarrow \underbrace{P_M(C_r = 1 \mid C_1 = c_1, \dots, C_{r-1} = c_{r-1})}_{\text{conditional click probability}}$   
3: Generate  $c_r$  from  $Bernoulli(P_r)$   
4: end for

Utility-based metrics

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

Effort-based metrics

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

Introduction Basic models Applications Advanced models Present Future Summary
Expected reciprocal rank

$$\begin{aligned} \mathsf{ERR} &= \sum_{r} \frac{1}{r} \cdot \mathsf{P}(\mathsf{S}_{r} = 1) \\ &= \sum_{r} \frac{1}{r} \cdot \mathsf{R}_{u_{r}q} \cdot \prod_{i=1}^{r-1} \left( \gamma \cdot (1 - \mathsf{R}_{u_{i}q}) \right) \end{aligned}$$

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Features	s for rank	ing				
Yande	<b>X</b> san francisco —	62 million answers	×	Search		
Web Images Video	San Francisco T sanfrancisco.travel v San Francisco is home time local. This is the pla	bit of everything. When		$\alpha_{u_1q}$		
Translate More	en.wikipedia.org > Sar San Francisco (/sæn fr	en'siskou/), officially the City and	yclopedia d County of <b>San Francisco</b> , is the and the only consolidated city-	$\alpha_{u_2q}$		
	wikitravel.org > en/Sar San Francisco is a maj	or city in California, the centerpie errain, Victorian architecture, sc	ace of the Bay Area, well-known for its enic beauty, summer fog, and great	$lpha_{\it U_3\it q}$		
	San Francisco C sanfrancisco.com * The job market may see opportunities can be fou	Chy to navigate these days	taurants, Nightlife, Real but employment and career cial District and Silicon Valley's	$\alpha_{u_4q}$		
	SF City and County of sfgov.org * SFGov Visitors Key Ser and supervisors. District	vices: SF Travel Resources 1	able of links to <b>San</b> Franisco districts	$\alpha_{\textit{u}_{5}\textit{q}}$		

۸

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

#### 1 Introduction

- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research

#### 7 Summary

Ilya Markov

Introduction

Basic models

Applications

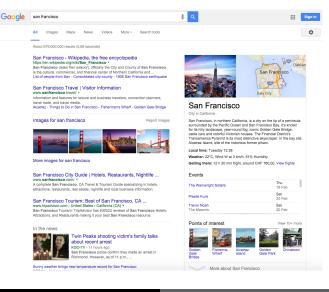
Advanced models

Present

Future

Summary

#### Aggregated SERPs



Ilya Markov

#### i.markov@uva.nl

#### Information Retrieval

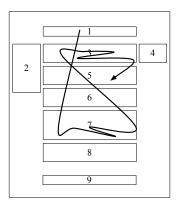
Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Search t	tasks					



Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Using fe	eatures					

- ID Feature description
- 1 Term frequency (TF) of body
- 2 TF of anchor
- 3 TF of title
- 4 TF of URL
- 5 TF of whole document
- 6 Inverse document frequency (IDF) of body
- 7 IDF of anchor
- 8 IDF of title
- 9 IDF of URL
- 10 IDF of whole document
- 11 TF\*IDF of body
- 12 TF\*IDF of anchor
- 13 TF\*IDF of title
- 14 TF\*IDF of URL
- 15 TF\*IDF of whole document
- 16 Document length (DL) of body
- 17 DL of anchor
- 18 DL of title
- 19 DL of URL
- 20 DL of whole document
- 21 BM25 of body
- 22 BM25 of anchor
- 23 BM25 of title
- 24 BM25 of URL
- 25 BM25 of whole document

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Beyond	clicks					



Picture taken from F. Diaz, R.W. White, G. Buscher, and D. Liebling. Robust models of mouse movement on dynamic web search results pages. In *CIKM*, 2013. ACM Press

Ilya Markov

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

#### 1 Introduction

- 2 Basic click models
- 3 Applications
- 4 Advanced models
- **5** Current developments
- 6 Future research

#### 7 Summary

Ilya Markov



#### Incorporating clicks, attention and satisfaction into a search engine result page evaluation model

Aleksandr Chuklin, Maarten de Rijke

Proceedings of CIKM 2016, Indianapolis, USA

 Introduction
 Basic models
 Applications
 Advanced models
 Present
 Future
 Summary

 A neural click model for web search

#### A neural click model for web search

Alexey Borisov, Ilya Markov, Maarten de Rijke, Pavel Serdyukov

Proceedings of WWW 2016, Montreal, Canada

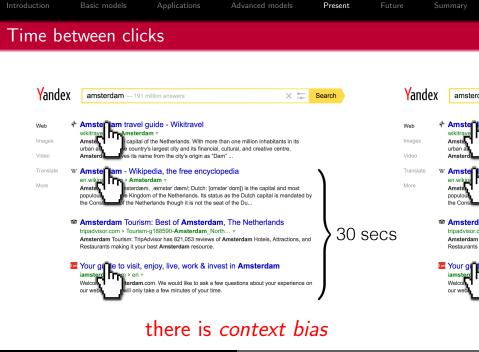
Advanced models Applications Advanced models Present Future Summary

### A context-aware time model for web search

Alexey Borisov, Ilya Markov, Maarten de Rijke, Pavel Serdyukov

Proceedings of SIGIR 2016, Pisa, Italy best student paper award Introduction Basic models Applications Advanced models Present Future Summary
Time between user actions

- Time between clicks
- Time to first click
- Time to last click
- Time between queries



Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Modelin	g time					

Average

 $Time("Amsterdam", "wikipedia.org") = rac{120+60+30}{3}$ 

Probability distribution

 $Time("Amsterdam", "wikipedia.org") \sim Gamma(\mathbf{k}, \mathbf{\theta})$ 

where  $(\mathbf{k}, \boldsymbol{\theta})$  are estimated from 120, 60, 30

## context bias is not modeled

Basic models Introduction Applications Advanced models Present Future Summarv

#### Context-aware time modeling



Introduction Basic models Applications Advanced models Present Future Summary
Context-aware time modeling

# $\begin{aligned} \textit{Time}(\textit{action},\textit{context}) &\sim \textit{Gamma}(\\ & & \textit{a}_k(\textit{ctx}) \cdot \textit{k}(\textit{act}) + \textit{b}_k(\textit{ctx}), \\ & & \textit{a}_\theta(\textit{ctx}) \cdot \theta(\textit{act}) + \textit{b}_\theta(\textit{ctx})) \end{aligned}$

Introduction Basic models Applications Advanced models Present Future Summary
Parameter estimation

 $\begin{aligned} \textit{Time}(\textit{action},\textit{context}) &\sim \textit{Gamma}(\\ & & \boldsymbol{a}_k(\textit{ctx}) \cdot \boldsymbol{k}(\textit{act}) + \boldsymbol{b}_k(\textit{ctx}), \\ & & \boldsymbol{a}_\theta(\textit{ctx}) \cdot \boldsymbol{\theta}(\textit{act}) + \boldsymbol{b}_\theta(\textit{ctx})) \end{aligned}$ 

- 1 Fix context-independent parameters
- 2 Optimize context-dependent parameters using neural networks
- ③ Fix context-dependent parameters
- ④ Optimize context-independent using gradient descent
- S Repeat until convergence

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Parame	ter estima	tion				

- We do not know the form of context-dependent parameters  $\implies$  neural networks
- We know the form of context-independent parameters (Gamma distribution) => direct optimization

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary	
Context							
		General					
	- (	ction) ´ erved time sinc	e previous action) previous action)	(0: no, 1: (0: no, 1: (0: undefir (0: undefir	yes) ned)		
		Q-action					
	BM25 (issue	h session erms in issued d query, previor ous query, issue	us query)	(0: no, 1: (0: undefir (0: undefir (0: undefir	ned) ned)		
		C-action					
	Is click on t	he 1 <sup>st</sup> position		(0: no, 1:	yes)		
	ls click on t	he 10 <sup>th</sup> positior	1	(0: no, 1:	yes)		



#### 3 months of log data from Yandex search engine

Time between actions	Max time	# Observations
Time-to-first-click	1 min	30, 747, 733
Time-between-clicks	5 min	6,317,834
Time-to-last-click	5 min	30, 446, 973
Time-from-abandoned-query	1 min	11, 523, 351

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Evaluati	on tasks					

#### Task1. Predict time between clicks

- Log-likelihood
- Root mean squared error (MSE)
- Task2. Rank results based on time between clicks  $\label{eq:list} \quad \ \ \, \text{nDCG} @\{1,3,5,10\}$

Present

Time model	Distribution	Log-likelihood	RMSE
Basic	exponential	-4.9219	60.73
	gamma	-4.9105	60.76
	Weibull	-4.9077	60.76
Context-aware	exponential	-4.8787	58.93
	gamma	-4.8556	58.98
	Weibull	-4.8504	58.94

Advanced models

Present

Summary

## Task 2. Ranking results

		NDCG			
Time model	Distribution	@1	<b>@</b> 3	@5	@10
Average	—	0.651	0.693	0.728	0.812
Context-aware	exponential gamma Weibull	0.668 0.675 0.671	0.710 0.715 0.709	0.743 0.748 0.745	0.820 0.822 0.821

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Other ti	imes					

- Time to first click
- Time to last click
- Time between queries

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Summa	ry					

- Removed context bias from time between actions
- Predicted user search interactions better (Task 1)
- Used the context-independent component for better document ranking (Task 2)

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

#### 1 Introduction

- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research

#### 7 Summary

- Keep on adding new variables not a good idea
- Parameter estimation
  - Efficiency
  - Online learning
- Other interactions and environments
  - Interactions beyond clicks
  - Devices beyond desktop computers

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Future r	research					

Model's output	Evaluation	Application
Conditional click probs	Log-likelihood	User simulation
Full click probs	Perplexity	Model-based metrics
Parameter values	Ranking evaluation	Ranking

- Why use intermediate evaluation?
  - Evaluate applications, not models
- Why maximize log-likelihood?
  - Optimize models for specific applications

Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						

#### 1 Introduction

- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research



Introduction	Basic models	Applications	Advanced models	Present	Future	Summary
Outline						



- 2 Basic click models
- 3 Applications
- 4 Advanced models
- 5 Current developments
- 6 Future research





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

