Beyond Ten Blue Links: Enabling User Click Modeling in Federated Web Search

Danqi Chen [*]
Institute for Interdisciplinary
Information Sciences,
Tsinghua University, China
cdq10131@gmail.com

Weizhu Chen
Microsoft Research Asia,
Beijing, China
wzchen@microsoft.com

Haixun Wang
Microsoft Research Asia,
Beijing, China
haixunw@microsoft.com

Zheng Chen
Microsoft Research Asia,
Beijing, China
zhengc@microsoft.com

Qiang Yang
Hong Kong University of
Science & Technolgy,
Hong Kong
qyang@cse.ust.hk

ABSTRACT

Click model has been positioned as an effective approach to interpret user click behavior in search engines. Existing advances in click models mostly focus on traditional Web search which contains only ten homogeneous Web HTML documents. However, in modern commercial search engines, more and more Web search results are federated from multiple sources and contain non-HTML results returned by other heterogeneous vertical engines, such as video or image search engines. In this paper, we study user click behavior in federated search results. In order to investigate this problem, we put forward an observation that user click behavior in federated search is highly different from that in traditional Web search, making it difficult to interpret using existing click models. Thus, we propose a novel federated click model (FCM) to interpret user click behavior in federated search. In particular, we introduce two new biases in FCM. The first indicates that users tend to be attracted by vertical results and their visual attention on them may increase the examination probability of other nearby web results. The other illustrates that user click behavior on vertical results may lead to more indication of relevance due to their presentation style in federated search. With these biases and an effective model to correct them, FCM is more accurate in characterizing user click behavior in federated search. Our extensive experimental results show that FCM can outperform other click models in interpreting user click behavior in federated search and achieve significant improvements in terms of both perplexity and log-likelihood.

[*] This work was done when the author was an intern at Microsoft Research Asia.

WSDM'12, February 8–12, 2012, Seattle, Washingtion, USA.
Copyright 2012 ACM 978-1-4503-0747-5/12/02 ...$10.00.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Measurement, Experimentation

Keywords

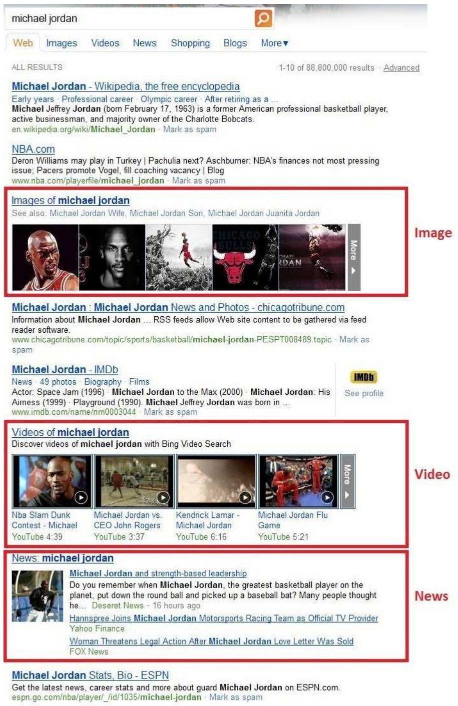
click model, federated search, log analysis

1. INTRODUCTION

Utilizing user click behavior in log data to understand user preference of search results is one of the most essential techniques for commercial search engines. Log data contains valuable information about user preference and can be collected at a low cost in most commercial search engines. This may in turn help search engines better entertain their users or deliver user-preferable advertisements. However, user click behavior in a commercial search engine may contain noise and biases. Many attempts have been made to address these challenges and most of them formalized this issue of learning an unbiased document relevance from user click data as a click model problem.

A well-known bias that needs to be corrected to learn an effective click model is position bias, where a document appearing in a higher position will attract more user clicks even it is not as relevant as other documents appearing in lower positions. As a result, the often used metric click-through rate (CTR) is not an exact measure of document relevance. Granka et al. [12] firstly observed this bias in their eyetracking experiments and a lot of research has been done since then with the goal of inferring an unbiased relevance. Richardson et al. [20] proposed to increase the relevance of documents in lower positions by a multiplicative factor. Craswell et al. [7] later formalized this idea as the examination hypothesis, which states that a document is clicked if and only if it is both examined and relevant. More recent work extended these methods to better interpret user click data in either organic Web search [5, 11, 13] or online advertising in sponsored search [26, 23, 21].

Despite of their successes, existing click models in organic Web search mostly focus on traditional Web search which contains only ten homogeneous HTML documents, which are often referred to as ten blue links in the prominent literature. In modern commercial search engines, more and more Web search results are federated from multiple sources and contain non-HTML results returned by other heterogeneous vertical engines. For example, the result of the query 'Michael Jordan' in Bing is shown in Figure 1. Besides the traditional HTML documents, it also contains results from vertical search engines such as *image, video* and *news*. It is obvious that the vertical results differ from traditional HTML documents due to the difference in presentation, layout, attractiveness, etc, which may in turn affect users' browsing behavior and their decision to perform clicks. However, previous click models are unable to address this issue since they assume the unique existence of ten blue links. As a result, modeling user click behavior in the federated search is an important but under-explored research problem.



ALL RESULTS 1-10 of 88,800,000 results · Advanced

Michael Jordan - Wikipedia, the free encyclopedia
Early years · Professional career · Olympic career · After retiring as a ...
Michael Jeffrey **Jordan** (born February 17, 1963) is a former American professional basketball player, active businessman, and majority owner of the Charlotte Bobcats.
en.wikipedia.org/wiki/**Michael_Jordan** · Mark as spam

NBA.com
Deron Williams may play in Turkey | Pachulia next? Aschburner: NBA's finances not most pressing issue; Pacers promote Vogel, fill coaching vacancy | Blog
www.nba.com/playerfile/**michael_jordan** · Mark as spam

Images of **michael jordan**
See also: Michael Jordan Wife, Michael Jordan Son, Michael Jordan Juanita Jordan

Michael Jordan : Michael Jordan News and Photos - chicagotribune.com
Information about **Michael Jordan** ... RSS feeds allow Web site content to be gathered via feed reader software.
www.chicagotribune.com/topic/sports/basketball/**michael-jordan**-PESPT008489.topic · Mark as spam

Michael Jordan - IMDb
News · 49 photos · Biography · Films
Actor: Space Jam (1996) · **Michael Jordan** to the Max (2000) · **Michael Jordan**: His Airness (1999) · Playground (1990). **Michael** Jeffrey **Jordan** was born in ...
www.imdb.com/name/nm0003044 · Mark as spam

Videos of **michael jordan**
Discover videos of **michael jordan** with Bing Video Search

Nba Slam Dunk Contest - Michael
YouTube 4:39

Michael Jordan vs. CEO John Rogers
YouTube 3:37

Kendrick Lamar - Michael Jordan
YouTube 6:16

Michael Jordan Flu Game
YouTube 5:21

News: **michael jordan**

Michael Jordan and strength-based leadership
Do you remember when **Michael Jordan**, the greatest basketball player on the planet, put down the round ball and picked up a baseball bat? Many people thought he... Deseret News · 16 hours ago

Hannspree Joins **Michael Jordan** Motorsports Racing Team as Official TV Provider
Yahoo Finance

Woman Threatens Legal Action After **Michael Jordan** Love Letter Was Sold
FOX News

Michael Jordan Stats, Bio - ESPN
Get the latest news, career stats and more about guard **Michael Jordan** on ESPN.com.
espn.go.com/nba/player/_/id/1035/**michael-jordan** · Mark as spam

Figure 1: Screenshot of the Bing SERP page showing vertical search results at different slots.

In this paper, we first study the user behavior in federated search and put forward an observation that user behavior in federated search is highly different from that in traditional Web search. To better characterize user behavior in federated search, we propose a novel Bayesian model called Federated Click Model (FCM), which introduces two new biases in order to capture the distinctive user behavior in federated search. The first illustrates that users tend to be attracted by vertical results and the visual attention on them will increase the examination probability of other nearby web results. This motivates us to reconsider the examination probability of each document in federated search and develop an attention model beyond the cascade hypothesis [7]. The other bias is due to the fact that vertial results have a special presentation style in federated search, user clicks are more highly correlated with the relevance of the results, which may impact the examination probability of the other Web documents. FCM takes these biases into account, leading to better characterizations of user click behavior in federated search. We conduct extensive experiments on a large-scale commercial dataset to evaluate the effectiveness of FCM in federated search. Experimental results demonstrate that FCM outperforms other existing click models and achieves significant improvements in terms of both perplexity and log-likelihood.

2. BACKGROUND & RELATED WORKS

In this section, we present the background and related works of click model, as well as works more closely related to federated search.

2.1 Click Model

We begin by stating some definitions and notations of click model. When a user submits a *query* q, the search engine returns a list of ranked documents as *search results*. A search session within the same query is called a *session*, denoted by s. The user *examines* some search results and *clicks* some or none of them. In our study, we only consider the first page of a session. We assume that the *search engine results page* (SERP) contains M documents, denoted by $\{d_1, d_2, \ldots, d_M\}$ where d_i is the document at position i from the top of the page. Usually, there are 10 traditional Web documents on the first page.

We use three binary random variables E_i, C_i and R_i to represent the examination, click and document relevance events of the document at position i.

- $E_i = 1$: the document at position i is examined.
- $C_i = 1$: the document at position i is clicked.
- $R_i = 1$: the document at position i is relevant.

Specifically, C_i is observable from the sessions while E_i and R_i are hidden random variables. We use the parameter r_{d_i} to represent the *document relevance*:

$$P(R_i = 1) = r_{d_i}. \tag{1}$$

Next, we introduce two important hypothesis: *examination hypothesis* and *cascade hypothesis*, which are the foundations of most existing click models.

Examination Hypothesis ([20, 7]) assumes that a document is clicked if and only if it is examined and relevant, which can be formulated as:

$$C_i = 1 \Longleftrightarrow E_i = 1, R_i = 1 \tag{2}$$

where E_i and R_i are independent. More precisely, we can represent the click-through rate as:

$$P(C_i = 1) = \underbrace{P(E_i = 1)}_{\text{position bias}} \underbrace{P(C_i = 1|E_i = 1)}_{\text{document relavance}} \tag{3}$$

where $P(R_i = 1) = P(C_i = 1 \mid E_i = 1)$ indicates the probability of click after examination. [20] assumes that the examination probability depends solely on position i, and thus $P(C_i = 1) = \lambda_i r_{d_i}$, in which λ_i models the position bias.

An extension of the examination hypothesis is the *user browsing model* (UBM) [11]. It assumes that the examination probability $P(E_i = 1) = \gamma_{i,i-l_i}$, which depends not only on position i, but also on the distance from the last clicked position $l_i = \max\{j : j < i, C_j = 1\}$. Thus

$$P(C_i = 1) = \gamma_{i,i-l_i} \times r_{d_i} \tag{4}$$

The *Bayesian browsing model* (BBM) [17] uses exactly the same model assumptions as UBM, but adopts a Bayesian algorithm for model inference.

Cascade Hypothesis [7] assumes that the user scans linearly from top to bottom of the search result page. Thus, a document is examined only if all previous documents are examined and the first document is always examined.

$$P(E_{i+1} = 1 \mid E_i = 0) = 0 \tag{5}$$
$$P(E_1 = 1) = 1 \tag{6}$$

Based on the hypothesis, the *cascade model* [7] constrains that a user will continue the examination until the first click, and then she abandons the whole search session:

$$P(C_i = 1 \mid E_i = 1) = r_{d_i} \tag{7}$$
$$P(E_{i+1} = 1 \mid E_i = 1, C_i) = 1 - C_i \tag{8}$$

The *dependent click model* (DCM) [14] generalizes the cascade model to allow multiple clicks:

$$P(E_{i+1} = 1 \mid E_i = 1, C_i = 1) = \lambda_i. \tag{9}$$

Another two important extensions to the cascade model are the *Click Chain Model* (CCM) [13] and the *Dynamic Bayesian Network Model* (DBN) [5]. CCM assumes that the probability to continue examination after a click depends on the relevance of the clicked document and ranges between two parameters α_2 and α_3:

$$P(E_{i+1} = 1 \mid E_i = 1, C_i = 0) = \alpha_1 \tag{10}$$
$$P(E_{i+1} = 1 \mid E_i = 1, C_i = 1) = \alpha_2(1 - r_{d_i}) + \alpha_3 r_{d_i} \tag{11}$$

DBN distinguishes the "perceived" relevance and the "actual" relevance and assumes that the actual relevance decides whether the user will examine the next document:

$$P(E_{i+1} = 1 \mid E_i = 1, C_i = 0) = \zeta \tag{12}$$
$$P(E_{i+1} = 1 \mid E_i = 1, C_i = 1) = \zeta(1 - s_{d_i}) \tag{13}$$

where parameter s_{d_i} is the actual relevance (or refered to as satisfaction) of document at position i.

Most recent works on click models incorporate more factors into user modeling: The *post-click click model* (PCC) [25] incorporates post-click behaviors (dwell time, whether a user has the next click, etc) into the click modeling; [15] argues that user clicks cannot be explained only by relevance and position bias, but also the user intents; The *session utility model* (SUM) [10] measures the relevance of a set of clicked documents as the probability that a user stops the session; The *joint relevance examination model* (JRE) [21] and *temporal click model* [23] capture the externality factor. [21] states that the relevance of a document is not a constant but affected by clicks in other positions and extends the UBM model by integrating a new parameter δ_{η_i} where η_i is the total number of clicks on positions other than i; [23] verifies the existence of externalities based on two advertisements and models their competition in click prediction. The *whole page click model* (WPC) [6] considers the search result page including the organic search and sponsored search as a whole to perform the CTR prediction; The *task-centric click model* (TCM) [24] characterizes user behavior related to a task as a collective whole; The *general click model* (GCM) [26] treats all relevance and examination effects as random variables and allows multiple biases.

The main divergence of our work with previous results is that we focus on understanding the user click behavior in federated search which contains vertical results. To the best of our knowledge, this is the first attempt to combine click model with federated search.

2.2 Federated Search

In traditional Web search engines, a ranked list of Web documents is retrieved in response to a user's query. This old-school paradigm becomes less effective since the information which the user is seeking is not always contained in a single document or category. Thus, to approach users' expectation, many search portals have assembled relevant contents from specific sub-collections, and placed them in an interface called *verticals*. [18] referred the method that integrates verticals into Web search results as *federated search* or *aggregated search*. Figure 1 introduces 3 major kinds of vertical results: image, video, and news, which are displayed in collections of single vertical documents embedded into the search engine results page (SERP).

This presentation of heterogeneous search results is believed to be promising and able to leverage user searching experience to higher levels. Previous studies mainly concentrate on several aspects:

- Most prior work [8, 16, 9, 3, 4] focuses on *vertical selection* — the task of determining, with a given query, which verticals, if any, should be presented alongside Web search results.
- *Vertical composition* is another important task, which is given multiple, already known to be relevant verticals, how one places them relative to Web search results [19, 2, 1].
- Furthermore, several studies have been done to investigate user preference behavior. For example, [22] investigates the effects of position and relevance on click-through behavior and finds a positive correlation between them.

Unlike prior works on federated search, we are working on a click model problem to automatically infer document relevance based on user click behavior. Besides the position bias in existing click models, we focus on illustrating the other biases introduced by the federated search and designing a new model to correct these biases to infer a more accurate document relevance. This is not covered by previous research.

3. USER BEHAVIOR ANALYSIS

In comparison with web results, vertical results contain media types which make user behaviors vary according to different entities. In this section, we investigate the impacts of verticals on the click-through rate (CTR) of web results and present a click log analysis to reveal the essential difference between vertical and web results. From the outcome of this analysis, we got an inspiration to build up an effective click model for federated search.

As a motivational experiment, we have collected a three-day click log dataset in June 2011 from Microsoft Bing search engine, and randomly sample $20,000,000$ query sessions with $4,276,432$ distinct queries. In this paper, we focus on 3 major verticals: image, video and news; only sessions of at least one image, video or news result are undertaken. We treat each vertical result as a single document and a vertical is clicked if one of its internal urls is clicked. Every kind of verticals displays at most once in the SERP; therefore, including 10 traditional web results, a SERP contains at most 13 results from top to bottom. Moreover, in our observed data, image, video and news verticals account for 14.23%, 31.04% and 54.73% respectively among all verticals. Figure 2 illustrates the distribution of positions where verticals are placed, and Table 1 indicates the average CTR of web and vertical documents over different positions. It is clear to see that the verticals are mainly placed in three positions: position 1 (top), position 4 (middle) and position 11 (bottom). Also, Table 1 shows that the average CTR of vertical documents is generally lower than that of the web documents, especially at the top 2 positions.

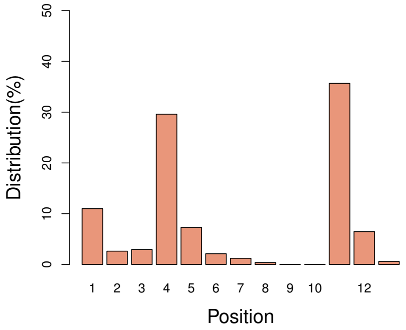


Figure 2: The distribution of positions where verticals are placed.

	@1	@2	@3	@4	@5
Web	53.71%	10.50%	5.09%	2.95%	2.27%
Vertical	19.09%	3.70%	8.35%	2.43%	1.08%

Table 1: The average click-through rate of web and vertical documents at position 1 - 5.

In the first experiment, we study the top 3 positions and explore the relationship between click patterns and types of the top results returned by the search engine. As receiving a high percentage of clicks, the top three results have a great positional advantage. Meanwhile, it is less likely that they are affected by other results underneath. We group sessions by the types of top three results, for example, “WVW” means that the top three results from top to bottom are from *web, vertical* and *web* respectively. For each group, we compute the probability distribution of 8 different click patterns $(000, 001, 010, 011, 100, 101, 110, 111)$ where 1 means a click while 0 means a skip. We discard sessions with the top results “VVV” and “VWV” which occur only few times and report the results in Figure 3. According to the results, we attain the following observations: (1) Comparing “WWW”, “WVV” with “VWW” and “VVW”, if the first impression is a web result, then with a high probability, users will click the first one and perform no clicks in lower positions. On the other hand, if the top results are verticals, users tend to scan from top to bottom until reaching the first web result, thereby the positional bias among the top results diminishes. (2) The difference between the distributions of “WWW” and “WWV” shows that, if the third result is a vertical, the probability of “100” decreases significantly, whereas “010” increases. The existence of a vertical result at position 3 decreases the position bias at above positions, i.e., for the case of “WWV”, users are more likely to examine the second results. (3) The group “WVW” has the highest probability of the click pattern “100” (71.50%). In the same manner, the existence of vertical will increase the CTR of the results in above positions while decrease that of the below results.

To further investigate the impact of verticals on the web documents, we perform a second experiment: In order to avoid the joint effects of several verticals, we only keep those sessions with exactly one vertical (more than 85.0% of the dataset has only one vertical). We group sessions by position and type of the vertical, then compute the average CTR of web documents at other positions. In addition, we retrieve the sessions without any verticals from the same three days’ log data and also compute the average CTR of web results at position 1 - 10. We only consider the sessions where the vertical is placed at position 1, 2, 3, 4, 5, 6 and 11 since position 7 - 10 only occupy 0.06% of the sessions in the whole dataset. The result is summarized in Table 2. The darker the color is, the more significantly the CTR increases. The average CTR at position 1 for traditional Web search is 57.27%. If there is an image vertical at position 3, the CTR of the first position decreases to 25.13%, but the CTR of the second position increases from 8.87% to 10.57%. In general, we observe that, except position 1, either image or video vertical has an obvious influence on the CTR of web documents, especially when the web documents are placed closely with the vertical. In particular, the CTR increases 100% $\sim$ 200% in many lower positions. On the other hand, there is no apparent increase when a news vertical is placed, this may be mainly attributed to the presentation style of news since it is actually still text-based document and the font sizes are similar as web results.

In the third experiment, we focus on users’ behaviors after clicking a document. We explore the probability of a click being the last click of the session in terms of the position and the type of the clicked document, since last click is believed to be informative and a good indication of user satisfaction [5]. Figure 4 illustrates that clicks on vertical results have significantly higher probability of being the last click except for position 1. For the exception of position 1 in Web, this may be driven by the high percentages of navigational queries like ‘microsoft.com’. For all the other positions, given a click, a vertical result is more likely to achieve user satisfaction. Specifically, image and video verticals have a higher probability than news verticals. This observation can also be explained by introducing the notions “perceived” relevance and “actual” relevance in [5]. The decision to click is made based merely on perceived relevance but the actual relevance decides whether the user is satisfied with the result, which in turn decides whether he will continue to examine following results. For a web document, users can only see a snippet including its title, url and a short description. However, for image, video and news documents,

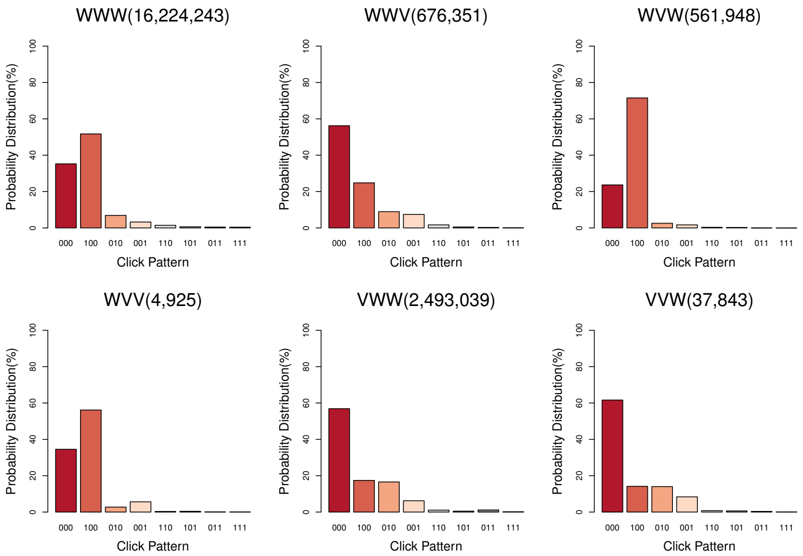


Figure 3: The probability distribution of click patterns over different results (top 3). The number in the parentheses refers to the number of sessions in the group.

search engine can offer an image thumbnail, a hover-to-play video thumbnail, even the news headline also delivers more information than Web. In other words, it is hard for users to examine a snippet and then judge whether it is really relevant or not. Rather, he can be aware of this from the thumbnail of an image or a video. Thus, the user will not continue to explore other results because his click on the image or the video most likely indicates his satisfaction.

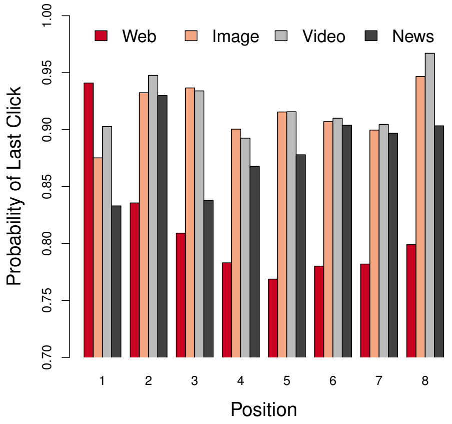


Figure 4: The probability of a click being the last click with respect to the position and the type of the clicked document.

In summary, vertical results and traditional web results mainly differ in three aspects: (1) Vertical results, especially image and video documents can attract more attentions and also make the surrounding web documents appear more prominent. (2) As a user clicks a vertical result, he is more likely to be satisfied with the document and then will not continue to examine other results. (3) For many users, they do not always seek vertical content, but may prefer the default web results instead. Therefore, web documents have a natural advantage to be trusted by users even in the federated search. (2) and (3) together explain why vertical results actually obtain more attentions but receive a lower click-through rate.

4. FEDERATED CLICK MODELS

Based on the observations and analysis in previous sections, in this section, we focus on developing an effective click model for federated search. We will propose two assumptions — *attention bias* and *exploration bias* to supervise the development of new click models.

4.1 Attention Model

According to the observations in Section 3, vertical results, especially image and video results tend to attract more attentions from users because of their visual presentation. As users' eyeballs are attracted by a vertical result, the other results close to it are more likely to receive more attention. We formalize this assumption as follows:

Assumption 1 (Attention Bias). *If there is a vertical placed in the SERP, users are more likely to examine the vertical as well as the web documents nearby. That is, vertical results play an extra role for attracting users' attention for other results around.*

To characterize the attention bias, we reconsider users' examination habits and present the new decision making flow in Figure 5. We use a binary random variable A representing whether there exists an attention bias in the current session s. We denote the probability of attention bias as $a(s)$ but it can vary based on the attention assumption design and incorporate many factors of verticals. If the session has no attention bias, the user will examine the document at position i with a probability ϕ_i where ϕ_i is the position bias in traditional click models. If there is an attention bias, the probability of examining a document will rise due to

Pos	1	2	3	4	5	6	7	8	9	10
No vertical										
avg	57.27	8.87	3.87	2.19	1.49	1.06	0.80	0.66	0.58	0.56
An image vertical placed at position 1, 2, 3, 4, 5, 6, 11										
1		21.58	10.07	6.19	4.04	3.01	2.31	1.89	1.58	1.38
2	46.76		6.04	2.89	1.60	1.04	0.79	0.65	0.46	0.39
3	25.13	10.57		5.75	3.96	2.60	1.86	1.48	1.16	1.02
4	26.29	11.61	7.34		4.18	3.53	2.37	1.98	1.67	1.46
5	36.47	12.54	7.66	4.97		3.12	2.46	1.98	1.54	1.38
6	28.16	14.38	10.14	6.38	4.55		3.05	2.45	1.99	1.79
11	47.28	12.10	6.63	3.99	2.78	1.94	1.45	1.10	0.88	0.70
A video vertical placed at position 1, 2, 3, 4, 5, 6, 11										
1		19.37	8.34	5.20	3.28	2.50	1.93	1.54	1.31	1.16
2	63.34		3.67	2.17	1.16	0.76	0.60	0.44	0.37	0.31
3	30.78	12.40		4.52	2.96	1.98	1.47	1.11	0.91	0.78
4	35.91	13.38	7.86		4.09	3.25	2.24	1.79	1.48	1.28
5	48.94	11.69	6.74	3.85		2.33	1.78	1.42	1.14	0.93
6	34.68	16.82	10.09	5.80	4.21		2.72	2.00	1.62	1.39
11	51.28	14.17	7.73	4.67	3.36	2.35	1.76	1.37	1.09	0.87
A news vertical placed at position 1, 2, 3, 4, 5, 6, 11										
1		23.82	8.04	4.55	2.53	1.97	1.46	1.08	0.91	0.80
2	72.55		2.88	1.32	0.69	0.48	0.36	0.28	0.20	0.19
3	58.18	6.53		2.12	0.87	0.88	0.47	0.33	0.31	0.41
4	60.50	7.57	3.35		1.06	0.99	0.70	0.52	0.43	0.36
5	74.18	4.58	1.71	0.80		0.48	0.32	0.24	0.20	0.17
6	68.31	4.98	2.18	1.00	0.86		0.35	0.33	0.25	0.23
11	54.80	6.72	3.03	1.63	1.15	0.75	0.55	0.42	0.34	0.26

Table 2: The CTR of web documents (%) at each position when there is no vertical or there is an image/video/news vertical.

the effect of verticals, which is written as $\phi_i + (1-\phi_i)\beta_{dist}$. $dist$ is an abbreviation for distance. β_{dist} is the parameter corresponding to the distance between the vertical result and the current result i. Generally, the closer the document from a vertical, the more attentions it receives from the user. Given the document is examined, following the *examination hypothesis*, the probability of click still depends on the document relevance r_{d_i}.

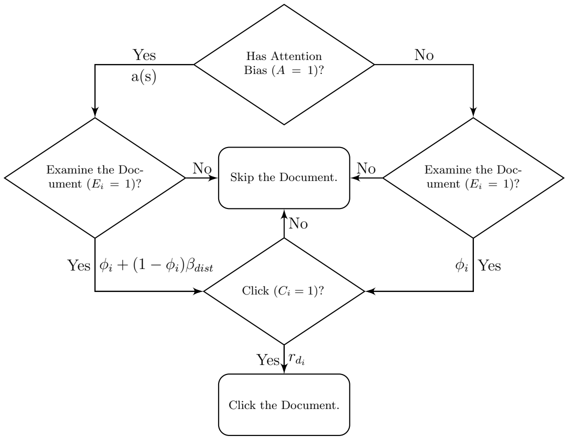

Has Attention Bias (A = 1)?
Yes
a(s)
No
Examine the Document (E_i = 1)?
No
Skip the Document.
No
Examine the Document (E_i = 1)?
No
Yes
$\phi_i + (1-\phi_i)\beta_{dist}$
ϕ_i
Yes
Click ($C_i = 1$)?
Yes
r_{d_i}
Click the Document.

Figure 5: The Attention Model.

To formalize the model, we may summarize it as the following conditional probabilities:

$$\begin{aligned}
P(A=1) &= a(s) && (14)\\
P(E_i=1 \mid A=0) &= \phi_i && (15)\\
P(E_i=1 \mid A=1) &= \phi_i + (1-\phi_i)\beta_{dist} && (16)\\
P(C_i=1 \mid E_i=0) &= 0 && (17)\\
P(C_i=1 \mid E_i=1) &= r_{d_i} && (18)
\end{aligned}$$

Since the majority of sessions in a search engine contain at most one vertical, for a sound illustration, we consider single vertical in the model specifications. On the other hand, it is natural to extend the single vertical to multiple verticals by treating $a(s)$ as the joint impacts of all verticals, and replacing β_{dist} with the distance from the closest vertical. Thus, we consider single vertical in the following model specifications to achieve simplicity.

In particular, we use β_{dist} to characterize the position effect, where $dist$ describes the distance between the web document at position i and the vertical. β_0 can be assumed as 1 with $\beta_0 = 1$, to assure the fact that if the session has an attention bias, the probability of examining the vertical itself is

$$\phi_i + (1-\phi_i) \times 1 = 1.$$

Furthermore, the degree of web documents influenced by a vertical may differ on the distance variable $dist$. If the position of the vertical is j while the current position for a Web document is i, we calculate $dist = j - i$ and each $dist$ has a corresponding parameter β_{dist}. Note that $dist$ may be a negative value and this is allowed in our model.

Another variable of the attention model is the parameter $a(s)$. We propose two different assumptions to develop $a(s)$ and call them as the *position attention model* and the *document attention model*.

4.1.1 Position Attention Model

The position attention model assumes that the attention bias of a session is decided only by the position of the vertical, where $a(s)$ is a position-specific parameter. We use pos_v to denote the position of the vertical in the SERP, then

$$P(A=1 \mid pos_v) = h_{pos_v}.$$

$\{h\}$ is a group of global parameters and can vary according to different type of verticals.

4.1.2 Document Attention Model

Another approach is to consider $a(s)$ as a document-specific parameter for each vertical. For different vertical documents, even they are placed in the same position, their attention bias values are different. Thus for each vertical document d, along with the relevance parameter r_d, we use another intrinsic parameter u_d to characterize the degree of the attention born with this document, thus

$$P(A=1 \mid d) = u_d$$

4.2 Exploration Model

According to the experimental findings in the third experiment of last Section, although vertical results receive a lower CTR, given a vertical result is clicked, it predominantly leads to more user satisfaction. Obviously, the increase of satisfaction may decrease the examination probability of other documents. We may formalize this as another assumption as follows:

Assumption 2 (Exploration Bias). *As a user clicks a vertical result, he is more likely to be satisfied with the result and stop the whole search session.*

We introduce a binary variable D to indicate whether the session has an exploration bias and C_v to indicate whether the user clicks a vertical result. If there is a click on a vertical result ($C_v = 1$), we use a parameter $e(s)$ to characterize the degree of the exploration bias of the session. We further assume that if the session has an exploration bias ($D = 1$), the user tends to skip web results. Furthermore, the probability of click ($C_i = 1$) still obeys the examination hypothesis. To distinguish web results and vertical results, we rely on a binary variable V_i, where $V_i = 1$ denotes a vertical document and $V_i = 0$ stands for a web document. The exploration model can be formalized as below:

$$\begin{aligned} P(D = 1 \mid C_v = 0) &= 0 & (19) \\ P(D = 1 \mid C_v = 1) &= e(s) & (20) \\ P(E_i = 1 \mid D = 1, V_i = 0) &= 0 & (21) \\ P(E_i = 1 \mid V_i = 1 \vee (D = 0, V_i = 0)) &= \phi_i & (22) \\ P(C_i = 1 \mid E_i = 0) &= 0 & (23) \\ P(C_i = 1 \mid E_i = 1) &= r_{d_i} & (24) \end{aligned}$$

where ϕ_i is the position bias in traditional click models. Similar to the attention model, we can also estimate the exploration bias $e(s)$ in two approaches — position-specific and document-specific — which are denoted by the *position exploration model* and the *document exploration model* respectively.

4.3 Joint Vertical Model

A natural generalization of the above two models is to integrate the attention model and the exploration model into one model. We call this the *joint vertical model* (JVM). We combine the examination probability as the product of the probability relying on the attention bias and the probability relying on the exploration bias. It is written as

$$P(E_i = 1 \mid A, D) = P(E_i = 1 \mid A) P(E_i = 1 \mid D) \qquad (25)$$

where $P(E_i = 1 \mid A)$ is given in Eq.(14), (15) and (16), $P(E_i = 1 \mid D)$ is given in Eq.(19), (20), (21) and (22). And the probability of click after examination still depends on the document relevance r_{d_i}.

In particular, each session has an attention bias $a(s)$ and an exploration bias $e(s)$ in the model. Following the previous models, we can define either of them in a position-specific or a document-specific approach.

The joint vertical model incorporates the attention bias and the exploration bias together, in which a placed vertical would increase the examination probability of surrounding web results but receive a low click-through rate. Meanwhile, if the vertical is clicked, the examination probability of all the web results will drop.

4.4 Inference

The proposed federated click models FCM can embrace the assumption of most existing click models dependent on the examination hypothesis. Till now, we assume ϕ_i as the position bias in traditional click models. Next, we may extend the FCM to embrace different assumption of ϕ_i based on different models. For the examination model in [20] we have,

$$\phi_i = \lambda_i.$$

For the user browsing model (UBM) [11],

$$\phi_i = \gamma_{i,i-l_i}$$

where l_i is the last clicked document. Even for the models relying on the cascade hypothesis, such as the dynamic bayesian network model (DBN) [5], it can be also extended naturally by combining Eq.(12),(13) and designed ϕ_i as:

$$\phi_i = \prod_{j<i} (\zeta(1 - s_{d_j})^{C_j})$$

We use the Expectation-Maximization (EM) algorithm to complete the inference step. The EM algorithm is used to find the maximum likelihood estimates of parameters, including the attention bias parameters $a(s)$, the exploration bias parameters $e(s)$, the distance parameters β_{dist}, the document relevance r_d and all parameters associated with ϕ (for example, in DBN model, the parameters s_d, γ are included). To perform EM iterations, in the E-Step, we compute the marginal posterior distribution of each hidden variable. The computation is performed based on the parameter values updated in the previous iteration. In the M-Step, all posterior probabilities associated with the same parameter are averaged to update the parameters.

We take an example to illustrate the inference where $\phi_i = \gamma_{i,i-l_i}$ as in UBM, $e(s) = \alpha_{pos_v}$, $a(s) = \eta_{pos_v}$ where $e(s), a(s)$ are only related to the position of the vertical pos_v. Next we show how to update $\{\eta\}$, and the other parameters can be derived similarly. As shown in Algorithm.1, at each iteration, we enumerate each document of each session and compute the posterior probability of η_{pos_v} using the parameters in the previous iteration, where pos_v is the position of the vertical in the current session. Finally, we update each η_p as the average posterior probabilities.

Algorithm 1 Inference

```
Iteration $k$:
for every position $p$ do
    $\eta_p^A \leftarrow 0, \eta_p^B \leftarrow 0$
for each document $d$ of each session do
    $pos_v \leftarrow$ the position of the vertical
    $i \leftarrow$ the position of $d$
    $l_i \leftarrow$ the position of last clicked document
    $dist \leftarrow$ the distance from the vertical
    if $d$ is a web result and $C_v = 1$ then
        $e \leftarrow 1 - \alpha_{pos_v}^{k-1}$ // $e$ is the exploration bias
    else
        $e \leftarrow 1$
    if $d$ is clicked then
        $\eta_{pos_v}^A \leftarrow \eta_{pos_v}^A + \frac{\eta_{pos_v}^{k-1} e(\beta_{dist}^{k-1} + (1-\beta_{dist}^{k-1})\gamma_{i,i-l_i}^{k-1})}{e(\gamma_{i,i-l_i}^{k-1} + (1-\gamma_{i,i-l_i}^{k-1})\eta_{pos_v}^{k-1}\beta_{dist}^{k-1})}$
    else
        $\eta_{pos_v}^A \leftarrow \eta_{pos_v}^A + \frac{\eta_{pos_v}^{k-1}(1-r_d^{k-1} e(\beta_{dist}^{k-1} + (1-\beta_{dist}^{k-1})\gamma_{i,i-l_i}^{k-1}))}{1-r_d^{k-1} e(\gamma_{i,i-l_i}^{k-1} + (1-\gamma_{i,i-l_i}^{k-1})\eta_{pos_v}^{k-1}\beta_{dist}^{k-1})}$
    $\eta_{pos_v}^B \leftarrow \eta_{pos_v}^B + 1$
for every position $p$ do
    $\eta_p^k \leftarrow \frac{\eta_p^A}{\eta_p^B}$
```

5. EXPERIMENTS

In this section, we compare the federated click model with two existing click models in Web search: UBM [11] and DBN [5]. For the FCM, there are three specific models. They are attention model, exploration model and joint vertical model (JVM). For each specific model, there are two assumptions: position-specific or document-specific. We use click perplexity and log-likelihood as metrics to measure the effectiveness of a click model.

5.1 Experimental Setup

The click logs used to train and evaluate click models are collected from Microsoft Bing search engine for a week in June 2011, which comprises $998,489$ randomly sampled queries. A session consists of an input query, a list of returned results on the search result page, each of which is either a web result or a vertical result, and a list of clicked positions. We keep those sessions with at least one vertical result (image, video or news). Only query sessions with at least one click are kept for performance evaluation. In order to prevent the whole dataset from becoming dominated by the extremely frequent queries, we allow each query at most 10^4 sessions. Finally, we collect nearly 27.5 million query sessions and 11.2 million web urls, 15.02% sessions have image impressions, 33.79% sessions have video impressions and 67.53% sessions have news impressions. The detailed information about the dataset is summarized in Table 3. For each query, we sort its search sessions by the timestamp when the query is sent to the search engine and split them into the training and testing sets at a ratio of $3:1$.

To evaluate the accuracy in click-through rate prediction, we trained different click models on the same training set. For the baseline models UBM and DBN, we use the inference algorithms introduced in the original papers. The training of our models follows the algorithm given in Section 4.4. Initially, all the relevance parameters are set to 0.2 and all the other parameters are set to 0.5. For either the attention model or the exploration model, we train the models from the two approaches — position-specific and document-specific for the attention bias $a(s)$ and the exploration bias $e(s)$. The iterative training of EM algorithms will run until all parameters converge. All the parameters related to the verticals are learnt according to the vertical types respectively. As the training is completed, we use the estimated parameters to predict the click-through rate for each document of every query sessions in the testing set. Utilizing the derived click probability, we evaluate the effectiveness of click models in terms of perplexity and log-likelihood.

5.2 Perplexity Evaluation

Perplexity is widely used to evaluate the accuracy of click-through rate prediction. It measures the accuracy for individual positions instead of the whole session. For a set of query sessions $s_1, \ldots, s_N$, if q_i^n is the probability of click at position i derived from the click model and C_i^n is the observed click event, the click perplexity at position i is:

$$p_i = 2^{-\frac{1}{N}(C_i^n \log_2 q_i^n + (1-C_i^n)\log_2(1-q_i^n))}.$$

The perplexity of the entire dataset is the average of p_i over all positions. A perfect click prediction will have a perplexity of 1 and a smaller number indicates better prediction accuracy. The improvement of perplexity value p_1 over p_2 is given by $(p_2 - p_1)/(p_2 - 1) \times 100\%$.

Table 4 illustrates the perplexity over positions from 1 through 8. It also reports the relative improvement of the federated click models over UBM. It shows that there is a clear ordering that JVM $>$ (attention model or exploration model) $>$ (UBM or DBN). More specifically, the UBM performs slightly better than DBN on perplexity, and all the federated models are significantly better than the UBM with more than 5% improvement of overall perplexity.

We can see that the two biases introduced in our federated click models can both learn a better accuracy in CTR prediction. Meanwhile, by combining the two biases together, JVM achieves a more significant improvement especially at the top 2 positions. It is clear to see that the JVM model has a 4.13%, 3.16% improvement at position 1 and position 2 respectively, and the overall improvement reaches 6.76%. In addition, if we compare either the position attention model with the document attention model, or the position exploration model with the document exploration model, we find that the position-specific models are slightly better than the document-specific models. This may be attributed to the skewness of the data since for a lot of low frequency queries, there is no enough data to learn an accurate document-specific parameter for each document. Thus in the following experiments, we use the position-specific approach to estimate $a(s)$ and $e(s)$.

5.3 Log-likelihood Evaluation

Another common metric to evaluate the effectiveness of click models is the Log-likelihood(LL). LL is computed as the average log probability of observed click events under the trained model. For a document d under a given query, if Pr is the derived probability of click, and C is the observed binary click event, then

$$LL = C\log_2 Pr + (1-C)\log_2(1-Pr)$$

The Log-likelihood of a dataset with a number of query sessions is measured as the average LL on individual documents. A larger LL indicates a better performance, and the optimal value is 0. The improvement of LL values l_1 over l_2 is computed as $(\exp(l_1 - l_2) - 1) \times 100\%$. We compute the average LL in terms of each types of documents, including web, image, video and news results. LL is often presented over the query frequency, which represents the number of sessions for a query in the training data. We separate the testing session based on the query frequency into 7 sets as Table 3. We then present the results of each set for both Web and different verticals in Figure 6.

It is also shown that all the federated click models outperform either UBM or DBN model for Web, especially for the low-frequency queries. Web results occupy about 80% of all the predictions thus it plays a major role in the overall prediction. The improvements on Set 1 and Set 2 in Web are more than 4% and 2% respectively for JVM, as compared with DBN. For the image and video results, attention model performs the best and achieves 2.04% and 2.26% improvements respectively.

6. CONCLUSIONS

In this paper, we have shown the user behavior difference between federated search and traditional Web search. We have demonstrated the necessity to model user specific behavior in federated search. Thereafter, we propose several

Table 3: The summary of the dataset in experiments. Image(#Queries, #Sessions) represents the number of queries and sessions with image impressions.

Set	Query Frequency	#Queries	#Sessions	Image		Video		News	
				#Queries	#Sessions	#Queries	#Sessions	#Queries	#Sessions
1	$1 \sim 10$	733,833	2,726,280	219,607	718,439	402,742	1,419,233	228,238	858,314
2	$10 \sim 10^{1.5}$	159,543	2,826,188	47,815	589,878	89,107	1,353,219	81,427	1,236,861
3	$10^{1.5} \sim 10^{2}$	67,239	3,650,037	23,059	634,601	37,272	1,490,827	45,188	2,085,483
4	$10^{2} \sim 10^{2.5}$	25,399	4,346,371	10,344	663,639	14,148	1,423,203	20,671	3,030,331
5	$10^{2.5} \sim 10^{3}$	8,489	4,550,372	3,684	611,990	4,802	1,234,566	7,561	3,571,209
6	$10^{3} \sim 10^{3.5}$	2,722	4,624,375	1,147	488,722	1,600	1,255,966	2,500	3,760,290
7	$10^{3.5} \sim 10^{4}$	882	4,738,227	377	418,657	500	1,102,033	832	4,002,307
Total		998,107	27,461,850	306,033	4,125,926	550,171	9,279,047	386,417	18,544,795

Table 4: The perplexity of comparison over ranking positions. "@n" represents the perplexity at position n.

	@1	@2	@3	@4	@5	@6	@7	@8	Overall
DBN	1.5487	1.4284	1.2758	1.1629	1.1351	1.1111	1.0902	1.0755	1.1660
UBM	1.5527	1.4253	1.2705	1.1624	1.1355	1.1118	1.0922	1.0789	1.1655
Attention(position)	1.5449	1.4197	1.2678	1.1569	1.1282	1.1027	1.0811	1.0660	1.1553
Impr.(vs. UBM)	1.42%	1.31%	0.99%	3.39%	5.36%	8.19%	12.10%	16.44%	6.11%
Attention(document)	1.5449	1.4203	1.2688	1.1571	1.1293	1.1027	1.0811	1.0659	1.1557
Impr.(vs. UBM)	1.41%	1.16%	0.62%	3.22%	4.54%	8.16%	12.07%	16.52%	5.87%
Exploration(position)	1.5469	1.4139	1.2720	1.1575	1.1280	1.1030	1.0819	1.0669	1.1559
Impr.(vs. UBM)	1.05%	2.66%	-0.56%	3.02%	5.48%	7.87%	11.19%	15.24%	5.79%
Exploration(document)	1.5518	1.4149	1.2720	1.1573	1.1278	1.1027	1.0816	1.0666	1.1564
Impr.(vs. UBM)	0.17%	2.44%	-0.56%	3.14%	5.65%	8.18%	11.53%	15.63%	5.47%
JVM	1.5299	1.4099	1.2683	1.1591	1.1286	1.1034	1.0819	1.0669	1.1543
Impr.(vs. UBM)	**4.13%**	**3.16%**	0.82%	2.01%	5.07%	7.54%	11.25%	15.30%	**6.76%**

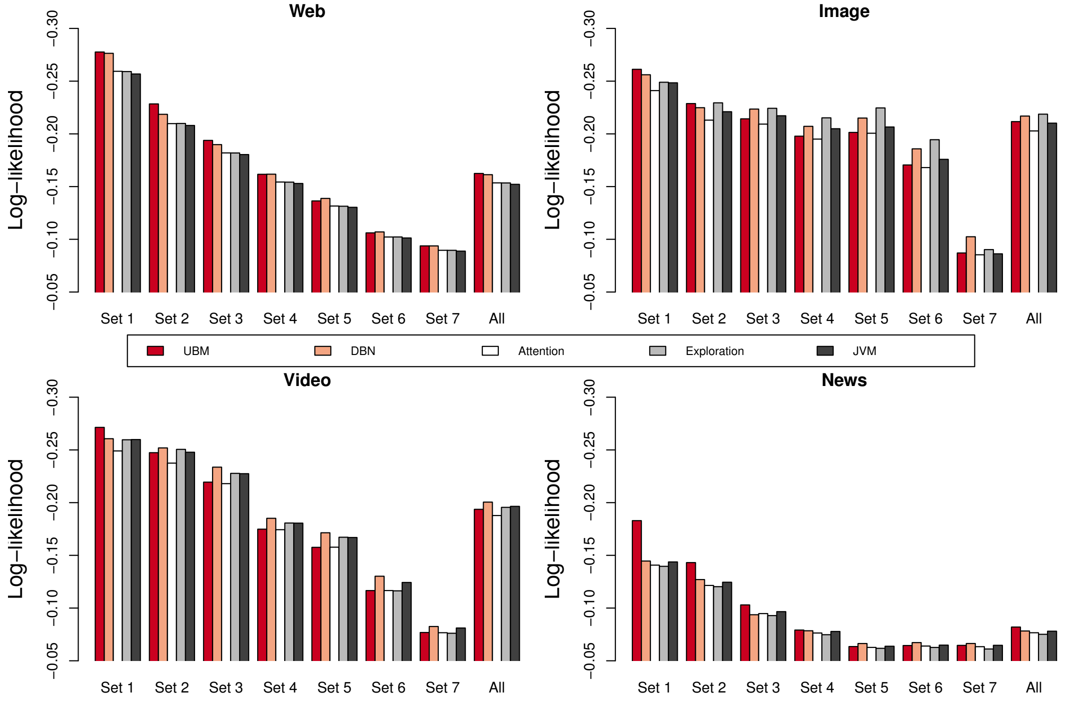


Figure 6: The log-likelihood of comparison over the web, image, video and news results.

federated click models to characterize user behavior as examining and clicking on vertical results in nowadays Web search. In particular, FCM has introduced two novel biases. The first indicates that users tend to be attracted by vertical results and their visual attention on them may increase the examination probability of other Web results. The other illustrates that user click behavior on vertical results may lead to more indication of relevance due to their presentation style in federated search. Taking these biases into consideration, we have designed FCM as a general model which makes it capable to embrace the assumptions in most existing click models. We have successfully extended it to embrace the assumptions of DBN and UBM, and designed an EM-based inference approach to make FCM capable of processing large-scale click data. Extensive experimental results have demonstrated that the proposed click models can better interpret user click behavior and achieve significant improvements in terms of both perplexity and log likelihood.

Thus, designing an effective click model to better understand user click behavior is critical but challenging for any commercial search engines. One of the challenges is the layout of the vertical results. Even with the same vertical, it will have multiple layouts for different queries. We plan to further study user click behavior by changing the positions in which the vertical results are placed. Some controlled experiments may be helpful on this, such as in an eye tracking environment. The other challenge is to better understand user intents when clicking on the vertical results. Even with the same query, different intents will lead to click on different vertical results. Both of these are our future works to explore.

7. ACKNOWLEDGMENTS

This work was supported in part by the National Basic Research Program of China Grant 2011CBA00300, 2011CBA00301, the National Natural Science Foundation of China Grant 61033001, 61061130540, 61073174.

Danqi Chen would like to thank Derek Hao Hu of Hong Kong University of Science Technology and three anonymous reviewers for their helpful comments.

8. REFERENCES

[1] J. Arguello, F. Diaz, and J. Callan. Learning to aggregate vertical results into web search results. In *CIKM*, 2011.

[2] J. Arguello, F. Diaz, J. Callan, and B. Carterette. A methodology for evaluating aggregated search results. In *ECIR*, 2011.

[3] J. Arguello, F. Diaz, J. Callan, and J.-F. Crespo. Sources of evidence for vertical selection. In *SIGIR*, 2009.

[4] J. Arguello, F. Diaz, and J.-F. Paiement. Vertical selection in the presence of unlabeled verticals. In *SIGIR*, 2010.

[5] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *WWW*, 2009.

[6] W. Chen, Z. Ji, S. Shen, and Q. Yang. A whole page click model to better interpret search engine click data. In *AAAI*, 2011.

[7] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *WSDM*, 2008.

[8] F. Diaz. Integration of news content into web results. In *WSDM*, 2009.

[9] F. Diaz and J. Arguello. Adaptation of offline vertical selection predictions in the presence of user feedback. In *SIGIR*, 2009.

[10] G. Dupret and C. Liao. A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine. In *WSDM*, 2010.

[11] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *SIGIR*, 2008.

[12] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *SIGIR*, 2004.

[13] F. Guo, C. Liu, A. Kannan, T. Minka, M. Taylor, Y.-M. Wang, and C. Faloutsos. Click chain model in web search. In *WWW*, 2009.

[14] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *WSDM*, 2009.

[15] B. Hu, Y. Zhang, W. Chen, G. Wang, and Q. Yang. Characterizing search intent diversity into click models. In *WWW*, 2011.

[16] X. Li, Y.-Y. Wang, and A. Acero. Learning query intent from regularized click graphs. In *SIGIR*, 2008.

[17] C. Liu, F. Guo, and C. Faloutsos. Bbm: bayesian browsing model from petabyte-scale data. In *KDD*, 2009.

[18] V. Murdock and M. Lalmas. Workshop on aggregated search. *SIGIR Forum*, 2008.

[19] A. K. Ponnuswami, K. Pattabiraman, Q. Wu, R. Gilad-Bachrach, and T. Kanungo. On composition of a federated web search result page: using online users to provide pairwise preference for heterogeneous verticals. In *WSDM*, 2011.

[20] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *WWW*, 2007.

[21] R. Srikant, S. Basu, N. Wang, and D. Pregibon. User browsing models: relevance versus examination. In *KDD*, 2010.

[22] S. Sushmita, H. Joho, M. Lalmas, and R. Villa. Factors affecting click-through behavior in aggregated search interfaces. In *CIKM*, 2010.

[23] W. Xu, E. Manavoglu, and E. Cantu-Paz. Temporal click model for sponsored search. In *SIGIR*, 2010.

[24] Y. Zhang, W. Chen, D. Wang, and Q. Yang. User-click modeling for understanding and predicting search-behavior. In *KDD*, 2011.

[25] F. Zhong, D. Wang, G. Wang, W. Chen, Y. Zhang, Z. Chen, and H. Wang. Incorporating post-click behaviors into a click model. In *SIGIR*, 2010.

[26] Z. A. Zhu, W. Chen, T. Minka, C. Zhu, and Z. Chen. A novel click model and its applications to online advertising. In *WSDM*, 2010.