Predicting Search User Examination with Visual Saliency

Yiqun Liu[†], Zeyang Liu[†], Ke Zhou[‡], Meng Wang[⋆], Huanbo Luan[†], Chao Wang[†], Min Zhang[†], Shaoping Ma[†],
[†]Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science & Technology, Tsinghua University, Beijing, China
[‡]Yahoo! Research, London, U.K.
[⋆]School of Computer and Information, HeFei University of Technology, Hefei, China
yiqunliu@tsinghua.edu.cn

ABSTRACT

Predicting users' examination of search results is one of the key concerns in Web search related studies. With more and more heterogeneous components federated into search engine result pages (SERPs), it becomes difficult for traditional position-based models to accurately predict users' actual examination patterns. Therefore, a number of prior works investigate the connection between examination and users' explicit interaction behaviors (e.g. click-through, mouse movement). Although these works gain much success in predicting users' examination behavior on SERPs, they require the collection of large scale user behavior data, which makes it impossible to predict examination behavior on newly-generated SERPs. To predict user examination on SERPs containing heterogenous components without user interaction information, we propose a new prediction model based on visual saliency map and page content features. Visual saliency, which is designed to measure the likelihood of a given area to attract human visual attention, is used to predict users' attention distribution on heterogenous search components. With an experimental search engine, we carefully design a user study in which users' examination behavior (eye movement) is recorded. Examination prediction results based on this collected data set demonstrate that visual saliency features significantly improve the performance of examination model in heterogeneous search environments. We also found that saliency features help predict internal examination behavior within vertical results.

Categories and Subject Descriptors

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

Keywords

User behavior analysis; Visual saliency; Eye tracking

1. INTRODUCTION

Web search has reached a level at which a good understanding of user interactions may significantly impact its quality. Among all kinds of user interactions, examination is an important one that are studied by many research works. Our understanding on how users allocate their limited attention to search engine result pages (SERPs) can contribute to improving search UI design, result ranking, performance evaluation, ad delivery and many other research issues in Web search. It also plays a central role in the *Examination Hypothesis* [9, 40], which is the basis of most search engine click model construction efforts [6, 8, 44].

One of the most frequently adopted information sources in search examination studies is the eye movement data collected by eye-tracking devices. According to the findings in cognitive psychological studies, vision appears to be blurred during saccades and new information is only acquired during eye fixations in the reading process [36]. Therefore, eye fixation sequence on SERP is usually adopted as a strong signal of search examination behavior according to the *strong eye-mind hypothesis* [27], that there is no appreciable lag between what is fixated on and what is processed.

Although eye tracking studies are able to offer rich detailed information about users' examination behaviors, the high cost and inconvenience of eye tracking devices limit the application of this methodology.[1] Therefore, some prior studies [15, 19] try to use mouse interaction information (e.g. click, movement and scroll) as a cheap surrogate to model and predict users' examination behaviors. These works reveal a strong correlation between eye fixation and mouse positions. However, these mouse-interaction-based researches also have their limitations. In these research works, users' mouse interaction data is required to predict search examination behaviors. It means that it is impossible for us to predict examination behavior on newly generated SERPs which are not shown to users yet. Firstly, considering the fact that there are a large number of long-tailed queries which are only submitted by one or few search users [38], modeling examination behaviors on their corresponding SERPs becomes rather difficult. Secondly, more and more heterogenous components are federated into SERPs and many of them contain highly dynamic information (e.g. news verticals). It makes it rather questionable whether examination behavior predicted based on a previous SERP can be adopted to the current one since the contents may have partially changed. Therefore, a more practical examination prediction method should rely on static (cold-start) information of SERPs and avoid the usage of user interaction information.

One of the key concerns in click model construction researches is to infer users' examination probabilities on search results. Therefore, most existing click models propose their assumptions in how users examine results on SERPs. Some of these assumptions (e.g. in

[1]Although some inexpensive eye-tracking solutions such as the eye tribe (https://theeyetribe.com) exist, they still require each user to equip one on the PC and calibrate each time before usage, which makes it impossible for large scale user behavior data collection.

SIGIR '16, July 17-21, 2016, Pisa, Italy

DOI: http://dx.doi.org/10.1145/2911451.2911517

Cascade model [9], DCM [14], DBN [6], UBM [12]) regard the results as homogeneous and take the position factor into consideration. Meanwhile, some recent proposed models also try to model users' behavior on heterogeneous SERPs (e.g. FCM [7], VCM [44]) and thus incorporate presentation style information as well. Since models that only consider position information may not perform well for SERPs that contain heterogeneous components [44], the assumptions in click models that can deal with vertical results should be adopted in practical examination prediction tasks. However, these click models also suffer from the same problem as mouse interaction based works: they cannot be adopted to previously unseen SERPs because they rely on users' click-through information to infer the examination behavior.

To shed light on the research question and propose an effective examination prediction method which relies only on static features (features that can be collected without user interaction), we propose to predict user examination with visual saliency information on SERPs. Visual saliency[2], as a measure of the likelihood of a location to attract human visual attention, has been widely adopted in the communities of Computer Vision (CV) and Human Computer Interaction (HCI). Many research efforts [22, 23] have revealed a strong connection between human attention and saliency map. Considering the existence of attractiveness bias in user behavior on SERPs with heterogeneous components [7, 29, 34, 44], saliency features may be suitable to model the influence of these components whose presentation styles are different from organic search results.

Take the two SERP samples in Figure 1 for example, the figures 1(a), 1(b) and 1(c) are corresponding to a SERP with only organic results while figures 1(d), 1(e) and 1(f) are corresponding to a SERP with an image vertical result ranked at the 5th position. The left, middle and right figures show users' eye fixation heatmaps (a, d), mouse hover positions (b, e) and the SERP images' visual saliency maps (c, f), respectively. From these figures we can see that the mouse hover positions correlate well with users' eye fixation positions, which validates previous findings in [11, 15, 19] that mouse movement data can serve as a surrogate for search users' fixation behaviors. We also noticed that the visual saliency maps are similar with eye fixation maps in certain parts of the SERPs. Particularly, the salient points in Figure 1(f) are quite similar with the eye fixation points in Figure 1(d), which are both located close to the image vertical result. Salient points in Figure 1(c) do not accord so well with eye fixation points in Figure 1(a) but they still provide some traces of the fixation points at lower-ranked results (e.g. the 4th and 5th results).

From Figure 1, we find that visual saliency information may provide some insights into users' examination behaviors, especially when mouse movement information (those recorded in Figure 1(b) and (e)) is absent. The salient points may not reflect well the position bias phenomena (i.e. the golden triangle) because top-ranked results do not necessarily contain visually salient components. However, it helps provide information on the modeling of attractiveness of search results, which is particularly important for SERPs with heterogeneous components [7, 34, 44]. It makes us believe that a prediction framework based on traditional position factors and the newly proposed visual saliency information may be a better way than existing solutions in modeling the examination behavior of search users.

To our best knowledge, we are among the first to adopt visual saliency information in predicting search examination behavior. A recent work [30] also propose to incorporate content salience into

[2]In this paper, we use the term *visual saliency* to describe a computational model of users' observed examination behavior.

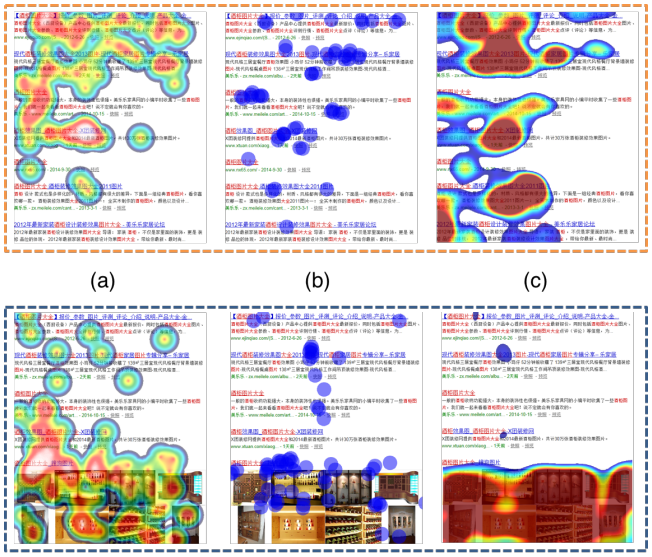

Figure 1: **Heatmaps of eye fixation (a, d), mouse hover (b, e) and visual saliency (c, f) on two search result pages (a, b, c: a SERP with only organic results; d, e, f: a SERP with both organic and image vertical results).**

predicting user attention on SERPs. Their proposed MICS method (Mixture of Interactions and Content Salience) is adopted to predict users' eye-fixation points on multiple popular types of Web content pages and gains much success. We believe that their work and our proposed method share a similar idea of incorporating salience information into examination models, but there are still many important differences. Firstly, their proposed method depends on user interaction behavior (e.g. mouse movement) in the prediction process. Therefore, it does not only rely on static features as our method and may also encounter the problem of inapplicability for newly-generated SERPs as other mouse-interaction-based methods do. Secondly, the MICS method in [30] uses page structure information (e.g. font size, image size, etc.) to calculate the salient points on Web pages, which means that it does not directly take the visual content information into consideration. Different from MICS, we are the first to use visual saliency maps derived from image content to predict users' examination behaviors. It makes it possible for us to take the influence of images with different color tones, edge densities and contrasts into consideration. Last but not least, our proposed method does not require the collection of extra user behavior data such as mouse movement information like in MICS. Experimental results in Section 5.5 also show that the visual saliency map extraction process can be quite efficient. Therefore, the proposed method is a relatively "cheap" solution for examination behavior prediction tasks.

Our contributions in this paper are three-fold: 1) We propose a novel examination prediction method that only utilizes the static information of SERPs, which makes it much more applicable to practical Web search environment than existing user-interaction-based solutions, especially for long-tailed queries. 2) By taking visual saliency map into consideration, the proposed method can be adopted to both homogeneous and heterogeneous search environment. 3) Besides page-level result examination behavior prediction, the proposed method can also model the internal examination behavior patterns of results within the vertical blocks.

The rest of the paper is organized as follows. Section 2 presents an overview of the related work. Then we introduce the experimen-

tal design for collecting user behavior data in Section 3. In Section 4 and 5 respectively, we describe the implementation of the saliency model, and discuss the results on using our model to predict search users' examination. We conclude the paper in Section 6.

2. RELATED WORK

Three lines of researches are closely related to the search user examination and attention prediction problem we describe in this paper: Web search eye-tracking studies, eye-mouse coordination, and visual saliency model.

2.1 Eye-tracking studies in Web search

The application of eye-tracking devices to Web search has received a considerable amount of attention from both academic and industrial researchers. Eye-tracking devices allow researchers to record users' real-time eye movement information, which helps better understand how users examine results on SERPs.

Granka et al. [13], Richardson et al. [40] and Joachims et al. [24, 25] used eye-tracking devices to analyze users' eye fixation distributions on SERPs and examination sequences throughout search tasks. Cutrell et al. [10] further investigated how users' eye movement behavior varies with different search intents. Recent work [34, 44] found that different result appearances might create different biases on eye movement behavior for both vertical and organic results on SERPs. Navalpakkam et al. [37] found that the flow of user attention on nonlinear page layouts is different from the widely believed top-down linear examination order of search results.

Based on these findings, a number of generative click models [6, 8, 9, 12, 44] have been constructed to model users' behavior during the search process. Most of these studies follow the strong eye-mind hypothesis [27] and regard eye fixation sequences to be the same as users' examination sequences.

2.2 Eye-Mouse Coordination in Web search

While eye movements during a search process can give us much insight into users' examination behavior, it is not applicable at large-scale in practice due to the high expense of eye-tracking devices. Therefore, many researchers turn to mouse movement information, which can be collected at large scale to simulate eye movements. Rodden et al. [41] identified multiple patterns of eye-mouse coordination, including the mouse following behavior in both vertical and horizonal directions while the eye inspected results. They also found a general correlation between eye and mouse positions, where the centers of the distribution of the eye/mouse distances are quite close to each other. Huang et al. [19] extended these findings by investigating variations in eye-mouse distances over time. They found that the distance between eye fixated points and cursors peaked approximately 600 ms after page loading and decreased over time. They also found that the mouse tended to be behind eye gaze by approximately 700 ms on average. These findings show that mouse movements may not be a good indicator for the current eye fixation position. However, in the whole session level it can help us identify which results are ever fixated by users.

Huang et al. [21] found correlations between result relevance and the cursor hovering behavior on the SERP. They incorporated mouse hover and scroll information as additional signals into click models to improve click prediction performance and gain promising results [20]. Guo et al. [16] analyzed the relationship between examination patterns and result relevance from post-click behaviors including cursor movements on landing pages. They constructed a predictive model to capture these patterns in order to improve search result ranking. Although mouse movements can be collected inexpensively in large-scale experiments, it is still difficult for commercial search engine to collect this kind of information in practical environment because it leads to extra burden for front-end servers. Considering the existence of the situations where users' behavior information is unavailable (e.g. on newly-generated pages), it is difficult for these models to predict users' behavior. Using static features of Web pages makes it possible to address this problem and predict users' examination without collecting users' interaction information.

2.3 Saliency-based Model

The idea of identifying salient items in search process has been investigated in a number of existing researches such as [2], in which "salient result" is defined as the one that users should examine or click in the next round of actions according to a theoretical framework (e.g. Search Economic Theory). However, in this paper, we focus on the concept of visual saliency of results on SERPs. Koch and Ullman [28] were among the first to introduce the concept of a saliency map, which is an explicit two-dimensional map that describes the saliency of objects or regions in the visual environment. Based on the idea of saliency map construction, Itti et al. [22] proposed an algorithmatic implementation to model saliency-based visual attention. They presented a bottom-up model to describe the pre-attentive selection mechanism, which works in three steps. Firstly, visual features (e.g. color, intensity and orientation) are extracted from the input image in the scene. Secondly, saliency features compete with each other in each feature map according to a certain strategy (e.g. winner-take-all). Finally, these different feature maps are combined at each location and summed into an unique saliency map. Inspired by the architecture of this saliency model, many saliency-based models have been proposed. For example, GBVS model [17], which is a complete bottom-up saliency model, creates feature maps applying Itti's method and perform their normalization and combination process with graph-based approaches. By extending from traditional saliency framework, some prior work [26, 31, 42] incorporated middle-level or high-level image features into attention models and performed a supervised training to predict eye fixations on images. The results of these work indicated that higher level features may be more effective in the attention prediction tasks.

Because the mechanism of selective visual attention may direct our gaze rapidly towards objects of interest in our visual field [23], saliency maps are usually applied to help analyze and predict users attention distributions. For example, Peters et al. [39] proposed an extension of saliency model to predict users' attention while they are playing video games. Carmi et al. [5] applied saliency models in dynamic scenes. Some existing studies also develop attention prediction models based on saliency information, such as [30]. However, to our best knowledge, the visual saliency model has not been adopted in predicting search examination behavior in Web search scenario. Our study indicates that saliency features can improve the performance of examination models and play an important role in predicting users' examination in Web search.

3. COLLECTING USER BEHAVIOR

3.1 Collection Procedure

To investigate the relationship between examination behavior and visual saliency features, we perform a user study to collect examination behavior on SERPs with the help of eye-tracking device (i.e. eye saccades and fixations on different areas of the SERPs). Considering that the presentation style of vertical result may have a strong effect on examination behavior [7, 34, 44], we implement an experimental search engine system to control the ranking position

Table 1: **Example search tasks in the user study**

Task Initial Query	Incorporated Vertical Result
Ancient Greek Architectural style	Textual
The 9^{th} zone (movie)	Encyclopedia
Nike basketball shoes	Image-only
iTunes download	Application
Ebola virus mutation	News

of both organic and vertical results.

With the experimental system, the user study is performed in the following steps. Firstly, we prepare two warm-up tasks to ensure that each participant is familiar with the experimental procedure and search system. In this step, each participant is told that his/her eye movements will be recorded while they are performing a few search tasks so that we can design better search systems. After the two warm-up tasks, participants are asked to go through calibration processes with the eye tracking device. Then they are instructed to finish the same set of 30 search tasks. All of the search tasks adopted in this study are selected from real-world commercial search logs so that they contain the practical users' search intention (some example tasks are shown in Table 1). To make sure that all participants see the same SERP in each search task, we provided a fixed initial query and its corresponding first result page from a popular commercial search engine (the same one which provides search logs) for each task.

Participants are allowed to click on any result link and visit the landing page for as long as they wish. The purpose of this design is to simulate the realistic search scenario in an experimental environment. At the end of the experiment, participants were required to provide some feedback about their search experiences and compensated with about US$10.

3.2 SERP Generation

In order to thoroughly investigate the influence of saliency features in practical search scenario, both organic-only and federated search pages were taken into consideration in our experiment. All the results of these pages, including both organic and vertical results, are crawled from the same commercial search engine and the original ranking of these organic results were preserved in the SERP generation process. The same protocol was also adopted in our previous work [34]. Since prior studies [1, 3, 7, 34, 44, 45] reveal that different factors of verticals lead to different behavior biases in heterogeneous search environment, we also take three aspects of federated search into account:

- Vertical type. We choose the same set of vertical result types as in some existing works [34, 44], including textual, encyclopedia, image, application-download and news verticals.
- Vertical position. Each vertical result is placed at the 1st, 3rd or 5th position of an SERP shown to participants (corresponding to the top, middle or bottom of the first viewport, respectively).
- Vertical relevance. For each search query, we collect both a relevant vertical and an irrelevant one from the search engine. The irrelevant vertical was collected by revising the original query to a related but different query so that it is actually not relevant to the user query. It is worth noting that these two verticals are all presented in the same layouts and presentation styles on SERPs.

Therefore, we generate six different federated SERPs (relevant or irrelevant vertical at 3 position options) per search task. Each federated SERP comprises one specific vertical and nine organic results. When a participant starts a certain task, one of the corresponding vertical results is randomly integrated into the SERP at the 1st, 3rd or 5th position. Following the pre-procedure steps above, we finally generated 180 (30 search tasks $\times$ 3 positions options $\times$ 2 vertical relevance) federated pages and 30 organic-only SERPs in total. To make sure that all tasks are completed with equal opportunities in each SERP condition, we used a Graeco-Latin square design [4, 18] to show tasks and conditions to participants. All the generated SERPs can be found in the public available dataset (see Section 5.1 for more details) to prompt reproductivity of the findings in the work.

4. EXAMINATION PREDICTION WITH VISUAL SALIENCY

With the information collected from the procedure described in Section 3, we aim to verify in this section the assumption that visual saliency information is useful in the examination prediction task. We start by describing the saliency model adopted in our study, and then present the features and our prediction models.

4.1 Visual Saliency Modeling on SERPs

Saliency map is an explicit two-dimensional map that encodes the saliency or conspicuity of objects in the visual scene. Most saliency models [17, 22, 23] were biologically inspired and based on a bottom-up computational model, which assumes that competition among neurons in saliency map causes a single winning location that corresponds to the next users' attention. Typically, there are three main steps in the generation of saliency models. Firstly, the input image is decomposed into multiple low-level visual features (e.g. color, intensity and orientation) and transferred to a set of static feature maps based on these pre-attentive features. After that, saliency models combine all the feature maps into a unique saliency map after neurons in each feature map compete for salience. This is the key step in the generation procedure and different competitive strategies may significantly affect the performance of saliency models. In the last step, models detect the most salient location and predict user next attend target based on winner-take-all network.

Considering the fact that vertical results are more likely to attract users' visual attention, we believe that the saliency model may also be appropriate to describe the implicit bias of user examination in search scenario, especially in federated search environment. In other words, saliency models can help capture the salient sections of pages and model the examination bias caused by verticals or other heterogeneous elements.

To predict users' examination behavior on SERPs, we selected one traditional and two state-of-the-art saliency models to generate the static saliency features. The traditional saliency model is presented by Itti et al. [22] in 2000, which model has been adopted by many works [23, 26] to predict the first few seconds of user attention on a wide range of images. Besides the traditional model, two state-of-the-art saliency models were also selected. The first one is the Graph-Based Visual Saliency (GBVS) model proposed in [17], which is one of the best performers in the evaluation benchmark of MIT300[3]. The second adopted state-of-the-art model is the TIP model proposed in [32]. This model is based on intermediate features between the low-level and the object-level features, and show promising results in multiple benchmarks. In our experiment, we deploy these three models to generate saliency maps of SERP pages (with color information) and extract saliency features from the corresponding saliency map. Considering the scrolling behavior may cause modification of salience distribution, we only focus on the

[3]http://saliency.mit.edu/results.html

examination behavior in the first viewport (containing first five or six search results of the SERP in general). It means that we construct static saliency maps based on the first viewport and predict users' examination in the first viewport of pages. Since most visual saliency models are designed to predict the first few fixation points of users, it is reasonable to focus on the first viewport. From the practical application's point of view, the prediction of examination behavior on the first viewport is also most important because users pay much attention on this set of results due to the existence of position bias [13, 24] (i.e. users tend to examine the search results from top to bottom of the ranking list while the probability of examination decreases dramatically as the ranking position increases).

4.2 Extraction of Saliency-based Features

The aim of our study is to develop a general model to predict user examination which only relies on static information of the SERPs. We mainly take two kinds of static features into account: content features and visual saliency features (shown in Table 2).

Similar to content feature group in the MICS method [30], we also concentrate on both layout information about results' positions, sizes of area, font sizes, and content information such as the number of child elements and area of images. These features are all regarded as content features in our experiment and many of them are believed to contain content saliency information according to [30].

Table 2: **Content and saliency features in examination models**

Feature Name	Feature Description
Content	Position of the given element
	Type of the given element
	Number of links and images
	Left, top, width, height
	Number of child elements
	Area of text and image
	Text space divided by the element's area
	Font size of the element's text
Saliency	Sum, mean and variance of the given element
	Histogram vector of the given element

To extract visual saliency features, we firstly calculate the salience value for each pixel in the target image using the saliency models as introduced in section 4.1. It is worth noting that different input images lead to completely different saliency maps. As mentioned, in our study, we take the first viewport of pages as the input images of saliency models. Then we compute the statistic information such as sum, mean and variance of each result block with the salience value of generated saliency maps. In addition, to further record the distribution of salient points in result blocks, we generated a histogram of salience values (in 5 equal bins) in the target images as features. The boundaries of each bin depend on the minimum and maximum salience values in the target image.

4.3 Prediction Models

We utilize machine learning models for the purpose of predicting examination behavior. Specifically, machine learning models consist of three types of variants: the learning target, the feature set used to represent the data and the learning model used for training, which we describe respectively as below.

4.3.1 Learning Targets

Considering the application of prediction results in practical Web search engines, we are mainly interested in predicting the examination behavior of search users in three scenarios:

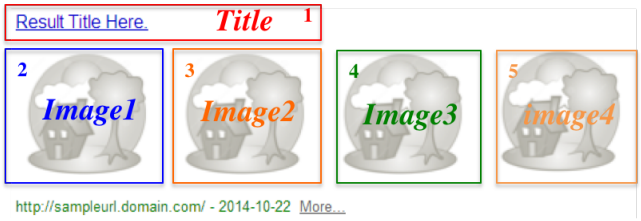

Figure 2: **An example vertical block which consists of a title part and four image components.**

- **Organic only SERP**: the search engine result page with only organic Web search results (i.e. traditional ten blue links);
- **SERP with a single vertical**: the search engine result page where a single vertical result block is embedded within the organic results (i.e. federated search page);
- **The vertical block**: the search result vertical block while an example of block for the image vertical is shown in Figure 2.

For each of these three scenarios, we mainly concentrate on two particular learning targets:

- **Attention Regression**: regress to the actual examination duration of each search results based on their total fixation duration obtained from eye-tracking devices;
- **Binary Examination**: classify the search results into two class: examined or not examined.

In order to obtain the fixation duration for the attention regression, we exclude all the saccade events and aggregate the eye fixation events for each search result in the SERP. To obtain the binary examination label (i.e. whether a given search result was examined or not), we deem all the search results that were fixated by users for at least 100ms as positive examples (examined) and treat the rest as negative examples. We tune various fixation thresholds (e.g. 100ms, 200ms, etc.) and found similar results in terms of the effectiveness ranking of models with different feature set. Therefore, in the rest of the paper, all our reported results on binary classification is based on the threshold of 100ms in users' fixations.

4.3.2 Features and Models

To thoroughly investigate the implicit effects of static (cold-start) features, we plan to compare different combinations of content and visual saliency features, which is shown in Table 3. Our baseline feature group is the one with Group ID = 1 which only contains content features (see section 4.2). This baseline can be regarded as an implementation of the cold-start MICS model [30] because they are based on the same static feature set.

Feature group SF only involves saliency features. Because result position plays an important role in users' examination processes, in feature group PSF we also consider the combination of position and visual saliency features. Feature group CSF is the union group of both content and visual saliency features, which contains all the static features of SERPs.

Note that we also report group 0 (PF), which only relies on the position of the given element (one of the content features) for comparison (but we do not use this as the baseline to compare against). Position bias is modeled in most existing click modeling efforts and we use this as a reference in order to track how we can improve over this. With respect to the vertical blocks, the ranking sequence is defined either from top to bottom or from left to right (e.g. image vertical block shown in Figure 2) in our experiment.

Table 3: **Feature groups in the examination models**

Group ID	Group Name	Features
0	PF	Position of the given element
1	CF	Content features
2	SF	Saliency features
3	PSF	Position and Saliency features
4	CSF	Content and Saliency features

To predict users' examination of results with the proposed feature groups, we experiment with five different learning methods that are widely-adopted in related studies: SVM, Logistic Regression (LR), Random Forrest (RF), Decision Tree (DT) and gradient boosting regression tree (GBRT). The implementation of scikit-learn toolkit[4] was adopted for these methods.

5. EXPERIMENTAL RESULTS

5.1 Experimental Setup

With the data collection procedure described in Section 3, we constructed an experimental dataset with users' eye movement information during the 30 search tasks. The dataset involves 35 undergraduate students (aged 18 to 25, mean = 18.8) as participants, all of whom were recruited from a wide range of majors of a university. Because of the calibration problems with the eye tracker, not all participants' eye movement data were available and 32 of them were finally taken into account. Therefore, we have altogether 32*30=960 search sessions in the dataset[5].

Considering that head-free eye trackers may make the collected interaction more natural and realistic, a Tobii X2-30 eye tracker was used to capture participants' eye movements and deployed the search system on a 17-inch LCD monitor whose resolution is 1366*768. Internet Explorer 11 browser was used to display the pages of search system. To identify users' examination behaviors, we detect fixations using built-in algorithms from Tobii Studio. In this dataset, users' fixation data on both organic only SERPs and SERPs with five different kinds of verticals (textual, encyclopedia, image-only, application, news) were collected and used as ground truth for the prediction experiments.

In our study, we not only systematically investigate the effects of saliency features in the prediction process on page level, but also focus on the prediction performance of internal examination within vertical blocks. For the prediction of internal examination, we need to identify which element is examined by users and extract the static features from these elements. To this end, we follow the methodology described in [30] by manually segmenting verticals into HTML DOM elements and selecting a subset of these elements as features. For the ranking of internal components within verticals, we assume that users prefer to examine vertical blocks from top to bottom and left to right. Therefore, we can label these segmented elements as a ranked list and take the rank of the elements as position features.

We use a variety of evaluation metrics to measure the performance of the trained model. Specifically, for binary classification (examined or not), we adopt the evaluation metrics of precision, recall, F-measure, accuracy and Matthews' correlation coefficient (MCC) score [35]. For attention regression, we mainly report in terms of the Pearson correlation and MSE (mean square error).

The prediction performance of the different methods is compared based on a five-fold cross validation and we report the average performance on the test folds.

[4]http://scikit-learn.org/stable/

[5]The dataset is available to download for research purposes at http://www.thuir.cn/group/~yqliu/publications/sigir2016Liu.zip.

5.2 Comparison of Machine Learning Models

We first compare the performance of different machine learning methods for the examination prediction task. Table 4 presents the comparison results of the five learning models as described in Sec.4.3.2. The baseline is corresponding to a naive majority judgment method in which we predict all results as examined ones (majority baseline). We just report a binary classification results due to limited space but the regression results follow a similar trend.

Table 4: **Examination prediction results of different learning methods in our dateset using both Content and Saliency Features (the saliency model of Itti [22] was adopted, bolded results are the best in corresponding columns, ** represents p-value < 0.01 compared with baseline)**

Model	Precision	Recall	F-measure	MCC	Accuracy
Baseline	0.618	1.000	0.764	0.000	0.618
GBRT	0.766**	0.783**	**0.775****	**0.401****	**0.719****
SVM	0.717**	**0.819****	0.764	0.316**	0.688**
LR	0.744**	0.798**	0.770**	0.364**	0.705**
RF	0.732**	0.791**	0.765	0.370**	0.701**
DT	**0.775****	0.693**	0.732**	0.360**	0.686**

From the results of Table 4, we can see that almost all learning frameworks outperform the majority baseline significantly in terms of F-measure, MCC and Accuracy. Especially, GBRT performs the best in most metrics. Therefore, we select GBRT as our machine learning model for training and prediction in the subsequent steps.

5.3 Comparison of Saliency Models

With the selected learning method, we focus on the contribution of visual saliency features generated by the three different models, i.e. Itti [22], GBVS [17] and TIP model [31] (see Section 4.1 for more details). To save space, we only report the results for the binary classification in Table 5 while we obtain similar results on the attention regression task.

Table 5: **Examination prediction results with visual saliency features generated by different saliency models**

	Organic only SERP		SERP with Vertical	
	F-measure	Accuracy	F-measure	Accuracy
Itti [22]	0.74	0.66	0.76	0.65
GBVS [17]	0.75	0.68	0.76	0.65
TIP [31]	0.73	0.64	0.77	0.65

From Table 5 we can see that the three visual saliency models do not lead to much differences in examination prediction performance. The TIP model works slightly better while being applied on SERPs with verticals but a little worse on the organic only SERPs. For the rest of the paper, we adopt TIP as the saliency feature generation model as we are more interested in the performance of heterogeneous search pages.

5.4 Examination Prediction Results

In this section, we report the examination prediction results of different feature groups on three different scenarios (as described in Section 4.3.1): page-level prediction on organic only SERPs, page-level prediction on SERPs with verticals and block-level prediction within vertical results. Due to space limitation, we only report two main evaluation metrics for each task: Pearson corre-

lation and MSE for the attention regression task, F-measure and Accuracy for the examination binary classification task.

5.4.1 Prediction on Organic only SERPs

We start by reporting the prediction results on organic only SERPs. From the example shown in Figure 1(a), 1(b) and 1(c), we assume that the position features may be good enough for the prediction task due to the existence of position bias on these SERPs. Visual saliency and content features may not be so useful in this circumstance because all results seem to be homogeneous ones (without the involvement of verticals). From the experimental results presented in Table 6, we verify our above assumptions with the following findings:

Firstly, the model that uses only the ranking position (PF) perform reasonably well while the content feature based baseline (CF) do not improve over the ranking position based baseline. This implies that on organic only SERPs, the font, area size and other content based features do not matter so much and users' attention is generally affected by the position factor.

Secondly, the saliency feature based model (SF) performs the worst, especially in the task of predicting attention duration (regression). This is as expected since on organic only SERPs, the results can be regarded as homogeneous ones and do not vary so much in terms of visual saliency. The only difference of saliency among the search results is that some results may contain more query terms (which are shown as bold and colored as red). In some cases, the top ranked results may have more exact matches and are therefore a bit more salient. However, this effect seems not so significant on the organic only SERPs.

Thirdly, when adding saliency and other content features in addition to the ranking position feature (CSF), we still can not improve over PF. This implies that it is difficult to outperform the position feature based model with the content and visual saliency features in the organic only search environment.

Table 6: **Prediction Performance on Organic only SERPs. Two-tailed t-test is performed to detect any significant changes against the prediction performance of content features (CF) (* represents p-value < 0.05 and ** represents p-value < 0.01)**

	Regression		Classification	
	Pearson	MSE	F-measure	Accuracy
PF	0.48	2.00E+06	0.77	0.72
CF	0.48	2.01E+06	0.77	0.72
SF	0.23**	2.40E+06*	0.73**	0.64**
PSF	0.48	2.04E+06	0.74**	0.68**
CSF	0.47	2.09E+06*	0.74**	0.68**

We also report the feature importance of the CSF model in Table 7. Due to space limitation again, we only report the prediction results of binary classification (examined or not). Not surprisingly, the ranking position (rank) and top result's position in pixel (top) are the two most important features in determining whether a given search result is examined or not. Interestingly, the saliency histogram information (e.g. saliency_hist4) that captures visual saliency on the pixel level can be helpful, while saliency_hist4 represents the fraction of level4 saliency pixels of this result against the whole SERP (1 means least salient and 5 means most salient). In addition, the saliency_var (saliency variance) feature, which quantifies the "saliency contrastness" of the result, is also useful. Note that not as we expected, we do not find that the area of the text, which quantifies the size of the each result, contribute much to the examination prediction task.

Table 7: **Feature Importance of the CSF Model while Predicting Examination on Organic only SERPs**

Top 1-5 Features		Top 6-10 Features	
Feature	**Weight**	**Feature**	**Weight**
rank	0.14	saliency_hist1	0.11
top	0.14	saliency_hist3	0.09
saliency_hist4	0.13	saliency_sum	0.05
saliency_var	0.13	saliency_ave	0.05
saliency_hist2	0.13	saliency_hist5	0.04

5.4.2 Prediction on SERPs with Verticals

As for the prediction task on SERPs with verticals, we believe that visual saliency features may be more effective since the SERPs are heterogeneous while vertical results maintain different presentation styles. Some vertical results may be more visually salient than others and this factor can affect users' examination process. The prediction performance of different feature groups on SERPs with verticals is shown in Table 8 and it also verifies our thoughts.

Table 8: **Model Performance on SERPs with Vertical. Two-tailed t-test is performed to detect any significant changes against the performance of CF model (* represents p-value < 0.05 and ** represents p-value < 0.01)**

	Regression		Classification	
	Pearson	MSE	F-measure	Accuracy
PF	0.36	2.02E+06**	0.76**	0.71
CF	0.43	1.87E+06	0.77	0.71
SF	0.39	1.94E+06**	0.77	0.65**
PSF	0.45	1.83E+06*	0.78**	0.72*
CSF	0.45	1.83E+06**	0.78**	0.72*

Firstly, position based features (PF) do not gain so promising results as on organic only SERPs, comparing all metrics of PF in Table 8 with those in Table 6. We can also find that the prediction performance of content features (CF) is better than that of PF and the difference is significant for MSE and F-measure. Comparing to the scenario of organic only SERPs, saliency based model (SF) performs better. It outperforms PF and is comparable to CF (although the performance differences in terms of MSE and Accuracy are still significant).

Secondly, especially in terms of MSE metric, adding visual saliency information to content based features (CSF) can further improve over the content based baseline (CF) and the difference is significant. For example, we found that PSF and CSF significantly (with paired two-tailed t-test) outperform CF with p-value < 0.05 and 0.01 respectively.

To investigate the performance of different features in the prediction process, we further show the feature importance of the CSF prediction model on SERPs with verticals in Table 9. We can observe that: again, the ranking position factors (rank and top) are very important in determining the examination behavior. Meanwhile, the "saliency contrastness" (saliency_var) and saliency histogram of medium levels (saliency_hist3 and saliency_hist2) become more important compared with the results in Table 7.

5.4.3 Prediction within Vertical Results

After investigating the page-level prediction results, we also look into the component-level prediction within vertical results, e.g. to predict which images are examined or paid much more attention in the image vertical block. From the results in Table 10 we can see that: Firstly, when predicting the examination of components within the vertical blocks, the ranking position features (PF) cannot

Table 9: **Feature Importance of the CSF Model while Predicting Examination on SERPs with Verticals**

Top 1-5 Features		Top 6-10 Features	
Feature	**Weight**	**Feature**	**Weight**
top	0.24	saliency_ave	0.08
saliency_var	0.16	saliency_sum	0.06
rank	0.14	saliency_hist4	0.04
saliency_hist3	0.09	saliency_hist1	0.04
saliency_hist2	0.08	saliency_hist5	0.03

compete with the content based features (CF). For all evaluation metrics, the results of PF model are worse than the CF model and the differences are significant with p<0.01.

Secondly, adding visual saliency features to position features (PSF) can help improve the prediction performance of PF. This improvement is reasonable considering the existence of multimedia components (e.g. images within the image vertical) and the heterogeneity nature (e.g. both image and text in the news vertical) of the vertical blocks. Thirdly, content feature based baseline (CF) performs the best, and significantly outperforms PF and SF, which show that the cold-start MICS model proposed in [30] is effective for predicting examination behavior within vertical results.

Table 10: **Model Performance on Internal Vertical Block.Two-tailed t-test is performed to detect any significant changes aganist the performance of CF model (* represents p-value < 0.05 and ** represents p-value < 0.01)**

	Regression		Classification	
	Pearson	MSE	F-measure	Accuracy
PF	0.22**	5.16E+05**	0.58**	0.65**
CF	0.39	4.65E+05	0.64	0.70
SF	0.33**	4.86E+05**	0.59**	0.67**
PSF	0.36*	4.81E+05**	0.62*	0.69**
CSF	0.38	4.70E+05	0.63*	0.69**

By examining the feature importance scores in Table 11, we can also obtain some insights why content feature based baseline performs so well. We can observe that the most important feature is the area_text, which represents the area size of a component within the vertical results. Not surprisingly, when users examine the items within the vertical blocks, the size of the area plays a vital part in attracting users' attentions. Meanwhile, the ranking position of the results (top) and the "saliency contrastness" (saliency_var) as well as saliency histogram of medium levels (saliency_hist3 and saliency_hist2) are also helpful in the prediction.

Table 11: **Feature Importance of the CSF Model while Predicting Examination within Vertical Results**

Top 1-5 Features		Top 6-10 Features	
Feature	**Weight**	**Feature**	**Weight**
area_text	0.19	saliency_hist4	0.06
top	0.15	saliency_sum	0.05
saliency_hist3	0.12	saliency_ave	0.05
saliency_var	0.07	left	0.05
saliency_hist2	0.06	saliency_hist1	0.05

5.5 Accuracy vs. Efficiency

From the experimental results in Sections 5.3 and 5.4, we find that visual saliency information helps improve prediction performance of users' examination behaviors. However, the feature groups that include saliency features (SF and CSF) require the processing of visual content information and may increase computational cost. To adopt this method in practical applications, we also want to find out whether the cost of saliency feature extraction is acceptable and how we can improve the efficiency if the cost is too much.

Table 12: **Prediction Model Performance (CSF) with Different Resize Resolution Ratios**

resolution	F-measure	Accuracy
82×46(6%)	0.77	0.72
95×54(7%)	0.77	0.71
109×61(8%)	0.77	0.71
122×69(9%)	0.77	0.71
136×77(10%)	0.77	0.71
272×154(20%)	0.77	0.71
544×307(40%)	0.78	0.72
816×461(60%)	0.78	0.72
1366×768(100%)	0.78	0.72

Considering that the TIP model outperforms the other two saliency models according to Table 5 in the federated SERPs, we just show the computational cost of TIP for generating saliency maps of federated SERPs in Figure 3. According to the figure, we can see that the average time for generating saliency map is around 44 seconds, which makes the extraction process almost impossible for practical applications when huge number of SERPs are processed. Therefore, we have to find a way to improve the efficiency of the feature extraction process.

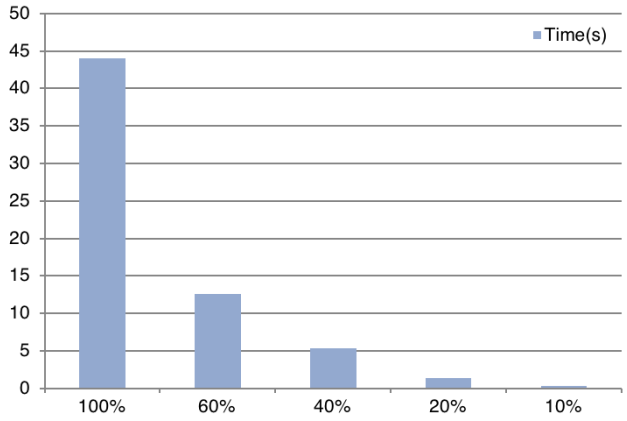

Figure 3: **Average computational cost (in Seconds) of generating saliency maps for the experimental data set.**

According to the selective attention theory [43] in cognitive psychology studies, human attention consists of two functionally independent, hierarchical stages: An early, pre-attentive stage that operates without capacity limitation and in parallel across the entire visual field, followed by a later, attentive limited-capacity stage that can deal with only one item (or at most a few items) at a time. Liu et. al. [33] show that in Web search examination process, a similar two-stage mechanism also applies. These existing research shows that search users will firstly allocate their attention in a skimming stage, in which no actual content is carefully read and understood. It reveals the possibility that low-resolution visual saliency features may also help in the examination prediction task.

To find out whether we can improve the efficiency of visual saliency model without loss of performance, we test the performance of visual saliency features extracted from SERP images with different resolutions. We resize the SERP image to different resolu-

tions with the Matlab imresize function using default interpolation method and antialiasing. After that, we test the computational cost of extraction for each resolution and corresponding performances for examination prediction. Since Table 8 shows that the CSF feature set performs best among all feature combinations, we test the classification results on SERPs with verticals using the CSF model. The experimental results are shown in Table 12 and Figure 3, respectively. Please be noted that we do not test the performance with resolution ratio less than 6% because that is the lowest allowed input for the imresize function.

According to the experimental results shown in Table 12, we can see that the model performance remain stable with different resize resolution ratios. Even when the image is resized to a resolution of 82×46, which is only 6% of the original image, the model performs almost as well as the original one. It indicates that although visual saliency features help improve examination prediction performance, there is no need to use the high-resolution image of SERPs. This phenomena is probably due to the fact that users just rely on the visual content to make a rough judgment of the possible interesting areas on SERP in the skimming stage of examination.

From Figure 3, we can find that the computational cost for resized images are greatly reduced. It only costs about 360 ms to generate the visual saliency map with a resized resolution rate of 10%. It makes the efficiency of the extraction process acceptable and shows that our proposed prediction model based on visual saliency features is applicable for practical Web search applications.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel examination model based on static information of SERPs, which has more practical applications in search scenario than existing user-interaction-based models. To our best knowledge, we are the first to use visual saliency maps in search scenario. With an in-depth study to analyze the impacts of saliency features in search environment, we demonstrate visual saliency features have a significant improvement on the performance of examination prediction. Importantly, saliency features, which contain the information of image content (e.g., color, brightness intensity), make it possible to predict user examination in complex search environment, especially in heterogeneous federated search. Without the information of user interaction, our model could offer good examination prediction only depending on the cold-start static information of pages. This could be quite valuable for the situations which lack users' behavior information, such as the evaluation of newly generated pages or in the new context (e.g. mobile). Further, we also confirm the positive effects of saliency features in the prediction of internal examination in vertical blocks. Our findings show that the proposed method can be also adopted to model the internal examination behavior patterns within vertical results. This may be particularly beneficial to the design of vertical layout.

Interesting directions for future work will include extending this work to construct click models based on static features of pages to improve search ranking performance. Moreover, inspired by the mechanism of inhibition of return in saliency map [22, 23] in cognitive psychology, we also plan to model the examination sequence depending on *dynamic* saliency maps in search environment.

7. ACKNOWLEDGEMENT

We thank Mr. Ming Liang and Dr. Xiaoling Hu for providing their visual saliency extraction implementations. This work is supported by Tsinghua University Initiative Scientific Research Program (2014Z21032), National Key Basic Research Program (2015CB358700), Tsinghua-Samsung Joint Laboratory for Intelligent Media Computing and Natural Science Foundation (61532011, 61472206) of China.

8. REFERENCES

[1] J. Arguello and R. Capra. The effects of vertical rank and border on aggregated search coherence and search behavior. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 539–548. ACM, 2014.

[2] L. Azzopardi and G. Zuccon. An analysis of theories of search and search behavior. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 81–90, 2015.

[3] J. Bar-Ilan, K. Keenoy, M. Levene, and E. Yaari. Presentation bias is significant in determining user preference for search results—a user study. *Journal of the American Society for Information Science and Technology*, 60(1):135–149, 2009.

[4] G. Buscher, S. T. Dumais, and E. Cutrell. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 42–49. ACM, 2010.

[5] R. Carmi and L. Itti. Visual causes versus correlates of attentional selection in dynamic scenes. *Vision research*, 46(26):4333–4345, 2006.

[6] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*, pages 1–10. ACM, 2009.

[7] D. Chen, W. Chen, H. Wang, Z. Chen, and Q. Yang. Beyond ten blue links: enabling user click modeling in federated web search. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 463–472.

[8] A. Chuklin, I. Markov, and M. d. Rijke. Click models for web search. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 7(3):1–115, 2015.

[9] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 87–94. ACM, 2008.

[10] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416. ACM, 2007.

[11] F. Diaz, R. White, D. Liebling, and G. Buscher. Robust models of mouse movement on dynamic web search results pages. In *Proceedings of the 22nd ACM international conference on Information and Knowledge Management (CIKM2013)*, pages 1451–1460.

[12] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 331–338. ACM, 2008.

[13] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479. ACM, 2004.

[14] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *Proceedings of the Second ACM*

International Conference on Web Search and Data Mining, pages 124–131. ACM, 2009.

[15] Q. Guo and E. Agichtein. Towards predicting web searcher gaze position from mouse movements. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3601–3606. ACM, 2010.

[16] Q. Guo and E. Agichtein. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proceedings of the 21st World Wide Web Conference 2012, WWW 2012, Lyon, France, April 16-20, 2012*, pages 569–578, 2012.

[17] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances in neural information processing systems*, pages 545–552, 2006.

[18] K. Hofmann, B. Mitra, F. Radlinski, and M. Shokouhi. An eye-tracking study of user interactions with query auto completion. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 549–558. ACM, 2014.

[19] J. Huang, R. White, and G. Buscher. User see, user point: gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1341–1350. ACM, 2012.

[20] J. Huang, R. W. White, G. Buscher, and K. Wang. Improving searcher models using mouse cursor activity. In *the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 195–204, 2012.

[21] J. Huang, R. W. White, and S. Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *the SIGCHI Conference on Human Factors in Computing Systems*, pages 1225–1234, 2011.

[22] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10):1489–1506, 2000.

[23] L. Itti and C. Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[24] T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161. ACM, 2005.

[25] T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems (TOIS)*, 25(2):7, 2007.

[26] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113, 2009.

[27] M. A. Just and P. A. Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[28] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

[29] D. Lagun and E. Agichtein. Effects of task and domain on searcher attention. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1087–1090. ACM, 2014.

[30] D. Lagun and E. Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 483–492. ACM, 2015.

[31] M. Liang and X. Hu. Feature selection in supervised saliency prediction. *Cybernetics, IEEE Transactions on*, 45(5):900–912, 2015.

[32] M. Liang and X. Hu. Predicting eye fixations with higher-level visual features. *Image Processing, IEEE Transactions on*, 24(3):1178–1189, 2015.

[33] Y. Liu, C. Wang, K. Zhou, J. Nie, M. Zhang, and S. Ma. From skimming to reading: A two-stage examination model for web search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 849–858. ACM, 2014.

[34] Z. Liu, Y. Liu, K. Zhou, M. Zhang, and S. Ma. Influence of vertical result in web search examination. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 193–202, 2015.

[35] B. W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[36] G. W. McConkie and K. Rayner. The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6):578–586, 1975.

[37] V. Navalpakkam, L. Jentzsch, R. Sayres, S. Ravi, A. Ahmed, and A. Smola. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd international conference on World Wide Web*, pages 953–964, 2013.

[38] S. Pandey, K. Punera, M. Fontoura, and V. Josifovski. Estimating advertisability of tail queries for sponsored search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 563–570. ACM, 2010.

[39] R. J. Peters and L. Itti. Applying computational tools to predict gaze direction in interactive visual environments. *ACM Transactions on Applied Perception (TAP)*, 5(2):9, 2008.

[40] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, pages 521–530. ACM, 2007.

[41] K. Rodden, X. Fu, A. Aula, and I. Spiro. Eye-mouse coordination patterns on web search results pages. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 2997–3002. ACM, 2008.

[42] T. Shi, M. Liang, and X. Hu. A reverse hierarchy model for predicting eye fixations. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2822–2829. IEEE, 2014.

[43] J. Theeuwes. Visual selective attention: A theoretical analysis. *Acta psychologica*, 83(2):93–154, 1993.

[44] C. Wang, Y. Liu, M. Zhang, S. Ma, M. Zheng, J. Qian, and K. Zhang. Incorporating vertical results into search click models. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 503–512. ACM, 2013.

[45] K. Zhou, R. Cummins, M. Lalmas, and J. M. Jose. Evaluating aggregated search pages. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 115–124, 2012.