Towards Measuring and Inferring User Interest from Gaze

Yixuan Li[1]*, Pingmei Xu[2], Dmitry Lagun[2], Vidhya Navalpakkam[2]
[1]Cornell University, Ithaca NY 14853
[2]Google, Mountain View, CA 94043
yli@cs.cornell.edu, {pingmeix,dlagun,vidhyan}@google.com

ABSTRACT

How can we reliably infer web users' interest and evaluate the content relevance when lacking active user interaction such as click behavior? In this paper, we investigate the relationship between mobile users' implicit interest inferred from attention metrics, such as eye gaze or viewport time, and explicit interest expressed by users. We present the first quantitative gaze tracking study using front-facing camera of mobile devices instead of specialized, expensive eye-tracking devices. We focus on multi-column digital media pages in Google Play Store that display 30+ items per page belonging to diverse categories. In such pages, we find significantly different distribution of gaze metrics on items that users rate as interesting vs. not. We leverage this insight by building a prediction model that is able to infer a user's interest ratings from the the non-click actions of the user. Our model is able to attain AUC of 90.32% in predicting user interest at an individual item level. In addition, our experiments on collection item re-ranking show how user gaze and viewport signals can be used to personalize item ranking on the collection page. Beyond understanding users' attention behavior in novel contexts such as multi-column digital media pages in Google Play Store, the findings in this study have implications for the design of a novel personalization and recommendation mechanism by (1) prioritizing items that are most likely of interest to users based on historical attention signals, and (2) prioritizing positions receiving significant portion of gaze attention.

Keywords

mobile phones; user attention; eye-tracking; personalization; user interest; inference

1. INTRODUCTION

For the past decade, web service providers have been relying mostly on explicit feedback, such as click signals, as indicators of content relevance to the users. The rapidly growing population of smartphone users has created the need for better understanding and harnessing users' behavior beyond just clicks. A common challenge is to reliably identify users' interest in scenarios when click data is sparse or unavailable. Several forms of implicit feedback including amount of scrolling and exit behavior have been found helpful in predicting explicit user feedback ratings [6]. However, these sources offer limited information about a users' focused attention, and the predictions are far from precise. In recent years, eye-tracking techniques have enabled researchers to reveal more subtle and fine-grained visual attention cues, and therefore address issues related to mobile users' behavior which cannot be otherwise addressed by using click signals.

We perform a controlled lab study to systematically investigate implicit gaze and viewport patterns captured from mobile devices as a source to measure users' attention and interest. We adopt a proprietary specialized eye-tracker that can estimate gaze position directly on a mobile phone, as users are performing their browsing tasks. Our experiments focus on Google Play Store, one of the largest digital distribution platforms containing rich contents of images and texts. Such digital channels can widely include, for example, app stores, recommendation sites (for restaurants, movies, music), shopping sites etc. Understanding users' attention and interest on digital distribution platforms can be crucial for improving user experience and satisfaction. An important feature of Play Store is the design of collection pages, which displays results grouped by pre-defined topics (e.g., popular games, trending new music etc.), or based on search queries specified by the user. The design and layout of Google Play Store is representative of many other online media and e-commerce and entertainment sites, thus making our findings applicable to many other settings.

Our work differs from previous eye-tracking research in two important ways. In terms of study design, unlike linear search results page (SERP), the digital collection pages seen on Play Store often display contents in multi-column layout. The nonlinear layout may add extra complexity, resulting in attention behavioral pattern that was previously not captured in conventional SERP. Second, in terms of methodology, our interest inference model is purely attention driven – without involving any contextual information of the display contents, nor explicit feedback such as clicks. In other words, our work examines the extent to which attention signals alone can be potent in revealing users' interest.

In this paper, we demonstrate how gaze metrics can be useful for understanding mobile users' browsing behavior on

*Work done while interning at Google.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054182

a wide range of tasks, including inferring user interest, and improving feed ranking and personalization in the context of mobile browsing. Specifically, our paper makes the following contributions:

- Presents the first quantitative gaze tracking study using front-facing camera of mobile phones instead of specialized, expensive eye-tracking devices. We use a proprietary gaze tracking system for end-to-end gaze estimation.

- Demonstrates statistically strong gaze attention bias towards left column on pages with double-column layout ($t(151) = 4.22$, $p < 0.001$). Such position effects (along the horizontal dimension) cannot be otherwise captured using previous viewport logging measurement [16].

- Identifies how users' attention pattern is affected by the vertical position of displayed content. Observes non-monotonic attention decay with rank position, with slight rebound towards the bottom of the page.

- Deploys an inference model to effectively predict users' interest on Play Store collection pages with accuracy 90.32% (AUC), using attention signals derived from gaze and viewport logging.

- Deploys a ranking model to personalize and improve feed relevance on Play Store collection pages, using attention related features.

We begin by surveying related work in eye tracking for user behavior on desktops and mobile devices in Section 2. We then describe our experiment and user study in Section 3, followed by the analysis of users' attention and interest on mobile phones in Section 4, Section 5 and Section 6. We further explore personalization of feed ranking with gaze in Section 7 and Section 8. We conclude with a discussion reviewing the findings and limitations of this study, along with suggestions for future work.

2. RELATED WORK

In recent years, scientists have been able to quantify and model users' attention behavior using eye-tracking techniques. Studies have been performed on various domains of application including web search [16, 13, 22], online news reading [17], smartphone app usage [24], recommendation systems [29] and ads quality [4] etc. Yang et al. [27] use eye-tracking in an experimental conjoint analysis to infer online consumer preference. Eye-tracking has also been used for understanding saliency of web pages [3, 26].

Our work focuses on studying user's attention pattern in a rather understudied realm – the mobile digital media sites. We take Google Play Store as an representative example of this type of platforms, and examine the gaze pattern on multi-column layout pages. Studies in the past have mostly focused on a linear page layout, e.g., search engine results page (SERP). In work that aligns more closely to our focus on nonlinear pages, Bota [2] studied the a novel interface of search results, and found that nonlinear composite results can positively impact search behavior in certain contexts. Navalpakkam et al. [23] studied the alignment between desktop mouse cursor and gaze position in search contexts with non-linear page layout. They show that the flow of user attention on nonlinear page layouts is different from the commonly seen top-down linear examination order of search results. Aside from mouse cursor, Claypool et al. used the amount of mouse scrolling time [6] as implicit feedback for making personalized recommendations. While these sources are readily available and useful on desktop, they offer limited information about a user's focused attention, and the predictions are far from precise.

Due to the expensiveness of commercial eye-tracking devices, there has been another line of research focusing on developing alternative measurement of mobile users' attention. Viewport (visible portion of a web page) has been demonstrated effective in approximating user attention. For example, viewport was used as an implicit feedback information to improve search result ranking for subsequent search queries [5], to help eliminate position bias in search result examination, detecting bad snippets and improving search result ranking [15]. Viewport time was also successfully used on mobile devices to infer user interest at the sub-document level [8]. Recently, Lagun et al. [16] found strong correlations between gaze duration and viewport duration on vertical search result, and that the average user attention is focused on the top half of the phone screen. [18] further employed viewport data to develop user engagement metrics that can measure user interaction during news reading and search results with ads. [18] studied how users' eye gaze (measured with viewport) and satisfaction are impacted by the presence of answer-like advertisements and their rich formats on SERP.

Our work contributes to the research field by demonstrating the use of gaze signals for reliably inferring users' interest, apart from their attention, in a novel context such as Google Play Store.

3. USER STUDY AND DATA COLLECTION

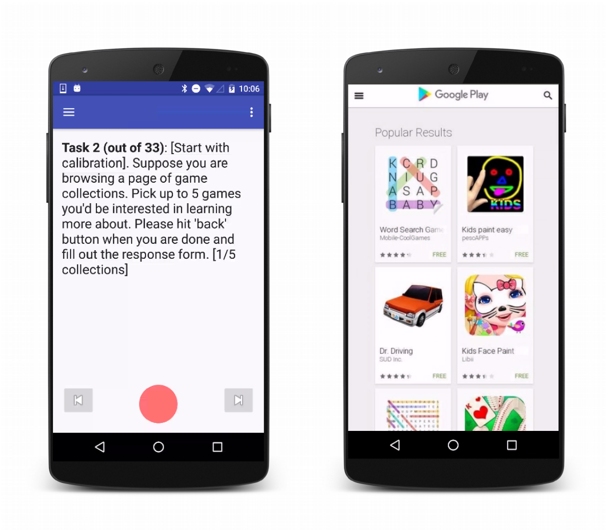
3.1 Participants

We recruited 36 participants with informed consent (20 male and 16 female), aged 18-60 (17 among them are > 30 years and 19 participants are $<= 30$ years), with various occupations and self-reported mobile search experience. Most of the participants had normal or corrected vision (e.g. wearing contact lenses) and were able to read from the mobile phone without wearing glasses.

3.2 Eye tracker

Our eye-tracker uses built-in camera on mobile devices for gaze estimation. This can be seen as a replacement for the expensive, specialized eye-tracker that has been commonly adopted for previous eye gaze user studies [7, 23, 16, 18].

The eye-tracker estimates gaze positions on the screen from eye-region images captured from the front-facing camera. In fact our system is very similar to [14, 28]. The training data is obtained through a *calibration* procedure. Duration the *calibration* each participant is asked to look at specific positions on the phone screen, marked with red circles. We use 13-point calibration system. A custom designed user study app displays a sequence of red circles one at a time, and at the same time records participants' appearance from the front facing camera. Each calibration session takes approximately 10 seconds. The image frames captured during the calibration stage together with corresponding lo-



(a) Task description (b) Collection page

Figure 1: **User study app interface. (a): The interface for task instruction. Participants are able to navigate through tasks by pressing "forward" and "backward" button. (b): Example of Google Play Store collection pages.**

cation of the red circles are used as training data to improve accuracy of the gaze estimation model.

For each participant, we fine tune gaze prediction model using data collected from all calibration sessions.[1] The fine tuned gaze estimation system has accuracy of 2.0° degree of visual angle.

3.3 Study Design

To better understand and validate the relationship between users' implicit eye gaze patterns and explicit interest, we design and conduct a lab study focused on Google Play Store. Our choice of performing experiments on Play Store is motivated by two main factors. First, a recent report [1] shows that mobile digital media time in the US significantly higher at 51% compared to desktop (42%). Play Store, as a popular digital distribution platform, contains rich contents of images and text. Second, the design and layout of Google Play Store is representative of many other mobile media and e-commerce sites. This makes our findings potentially applicable to many other settings.

Specifically, we simulate experiments on Play Store *collection pages*, which refer to pages displaying results grouped by pre-defined topics (e.g., popular games, trending new music etc.), or based on search queries specified by the user. The layout of a digital collection page can be multi-column, depending on the screen size and configuration of a phone. Our Nexus 5 phone used in this study displays results in a double-column layout, as illustrated in Figure 1 (b).

The study uses the following protocol. The experiments begin by calibrating eye gaze of each participant. After calibration, participants are presented with a task description page on the user study app. As shown in Figure 1, the description page instructs participants to freely browse a collection page, and mentally pick up to 5 items that they find interesting. Once finishing reading the description, participants can tap the red button to start the task (this also triggers video recording in the background). Participants will then be directed to a pre-defined collection page containing the thumbnails of 32 different games on Play Store. Each game collection page is generated by randomly sampling 8 games from each of the following four categories: *word games, racing games, board games, kids painting games.* These four categories were selected to capture the diversity within the gaming category. The order of games on each collection page is randomized to avoid potential effects of position bias.

Upon finishing browsing, participants can navigate back to the study home page by using the "Back" button on the phone. Participants are then asked to complete the post-task by check marking the games that they found interesting during the browsing session, thus revealing the users' explicit preferences or interest. This is a separate procedure, which ensures that the selection action does not interfere with the browsing process.

In addition to item-level interest, we collect users' interest information at the coarse categorical level. Each participant is asked to rate their preference for each of the four pre-defined game categories on a 10 point likert scale – 1 being not interested at all and 10 being completely interested.

There are in total 5 pre-designed mocks of Play Store game collection pages, each of which is generated using the same procedure described above. Each participant is asked to repeat the same "*calibration – browsing – selecting and rating*" process for all these 5 different pages[2].

3.4 Logging and Post-processing

All the viewport events (on the phone screen end) are buffered and subsequently sent with an HTTP request to a user study server which stores data for post analysis. Such instrumentation allows us to join stimuli data with gaze position time series and reconstruct what the user saw on the screen at any point of time.

4. ATTENTION METRICS

4.1 Gaze

We derive gaze-based attention metrics using our specialized eye-tracker. We denote $(y_h^{(j)}, y_v^{(j)}, t^{(j)})$ the estimated horizontal and vertical phone screen coordinates at timestamp $t^{(j)}$. Given the bounding box B of any area of interest (AOI) on a page, we compute gaze metrics as follows:

- **Gaze Dwell Time**: Amount of gaze time (in seconds) a user spends viewing the AOI.

$$\mathcal{T}_{\text{gaze}}(B) = \sum_{(y_h^{(j)}, y_v^{(j)}) \in B} t^{(j+1)} - t^{(j)}$$

- **Gaze Dwell Fraction**: The percentage of gaze time

[1]To ensure the accuracy of gaze estimation is sufficient, the calibration process we performed calibration multiple times during the user study. We require participants to perform calibration when the head pose or position substantially changed.

[2]The design choice of presenting 5 collection pages is made so that we can have adequate within-subject browsing data, without overwhelming the participant.

a user spends viewing the AOI.

$$\%\mathcal{T}_{\text{gaze}}(B) = \frac{\sum_{(y_h^{(j)}, y_v^{(j)}) \in B} t^{(j+1)} - t^{(j)}}{\text{Time on Page}}$$

- **Gaze Time to First Visit**: The first timestamp at which a gaze comes into the AOI.

$$\mathcal{T}_{\text{gaze\_tfv}}(B) = \min_{(y_h^{(j)}, y_v^{(j)}) \in B} t^{(j)}$$

4.2 Viewport

In addition to gaze, we also adopt viewport metrics proposed in previous studies [16, 17] for attention estimation. Viewport logging works by recording the portion of the web page visible on the screen at any given time, as well as bounding boxes of all displayed contents shown on the page. Compared to gaze, viewport is a less fine-grained attention metric since it estimates gaze by assigning viewing time proportional to the size of given AOI. We denote $\mathcal{V}^{(j)}$ the bounding box of phone screen viewport at timestamp $t^{(j)}$. Given the bounding box B of any area of interest (AOI) on page, we compute viewport metrics as follow:

- **Viewport Time**: The visible time of an AOI at a given viewport position on the phone screen.

$$\mathcal{T}_{\text{viewport}}(B) = \sum_j \frac{\mathcal{V}^{(j)} \cap B}{\mathcal{V}^{(j)}} \cdot (t^{(j+1)} - t^{(j)})$$

- **Viewport Dwell Fraction**: The percentage of time a user spends viewing an AOI at a given viewport position.

$$\%\mathcal{T}_{\text{viewport}}(B) = \frac{\mathcal{T}_{\text{viewport}}(B)}{\text{Time on Page}}$$

- **Viewport Time to First Visit**:

$$\mathcal{T}_{\text{viewport\_tfv}}(B) = \min_{\{\mathcal{V}^{(j)} \cap B\} \neq \varnothing} t^{(j)}$$

5. GAZE REVEALS POSITION EFFECT

The position effect on user attention has been extensively studied in the context of search, both on desktops [7] and mobile phones [16]. Different from conventional, vertical search results page (SERP), the grid layout on digital media platforms adds further complexity – causing position bias both in the horizontal and vertical direction. In this section, we examine how user attention is affected by position on grid layout pages.

5.1 Horizontal Position

To study the horizontal position effect, we divide the entire collection page into two AOIs: left column vs. right column, and derive attention metrics correspondingly. The left panel in Figure 2 shows box-plot of gaze fraction (%) on the left column and right column, respectively. Participants spend on average $52.91 \pm 1.50\%$ of fractional time on the left column, as opposed to $43.92 \pm 1.50\%$ of fractional time on the right column. The kernel density estimation of data distribution is visualized in the right panel of Figure 2. A two-sample t-test shows significant gaze time (%) bias towards the left column on collection pages ($t(151) = 4.22$, $^{***}p < 0.001$).

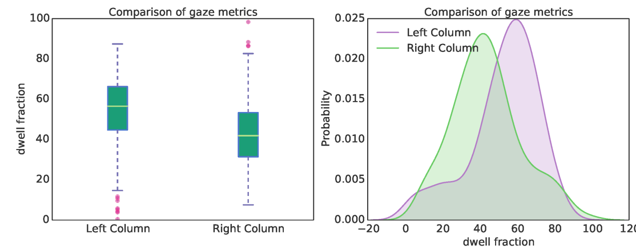


Figure 2: **Gaze dwell fraction between left and right columns on Play Store collection pages are statistically significantly different with highly non-overlapping 95% confidence intervals.**

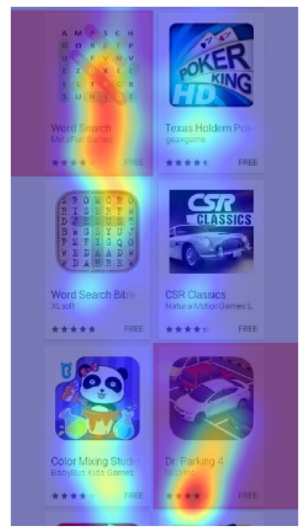


Figure 3: **Example heatmap of gaze on PlayStore game collection page. Note that the gaze hotspots are focused on items that a user reported later as interesting (the red highlights are for visualization purposes only, and were not part of the display to the user).**

It is worth noting that such position bias cannot be captured using previous viewport logging measurement. When measured with viewport metrics, the distribution of dwell fraction (%) between left and right columns are indistinguishable. This is due to that viewport metric has the shortcoming of coarsely assigning equal viewing likelihood on both sides, without being able to differentiate in between. Unlike viewport metrics, we see a clear difference in gaze patterns between the left and right columns.

5.2 Vertical Position

We also study the position effect in the vertical direction. The top panels in Figure 4 show the mean and variance measured by different gaze metrics (dwell time, dwell fraction and time to first visit) on AOIs of game thumbnails as a function of vertical position (#row). The statistics at any given position is based on data points from all 36 participants × 5 collection pages. The x-axis varies from 1 to 16 since there are 32 games on each collection page, with 2 displayed side by side per row. We plot in blue the metrics for AOIs on the left column, and green for the AOIs on the right column, respectively. The curves are fitted using polynomial regression with confidence interval of 95%.

A first salient observation from the top-left and top-middle

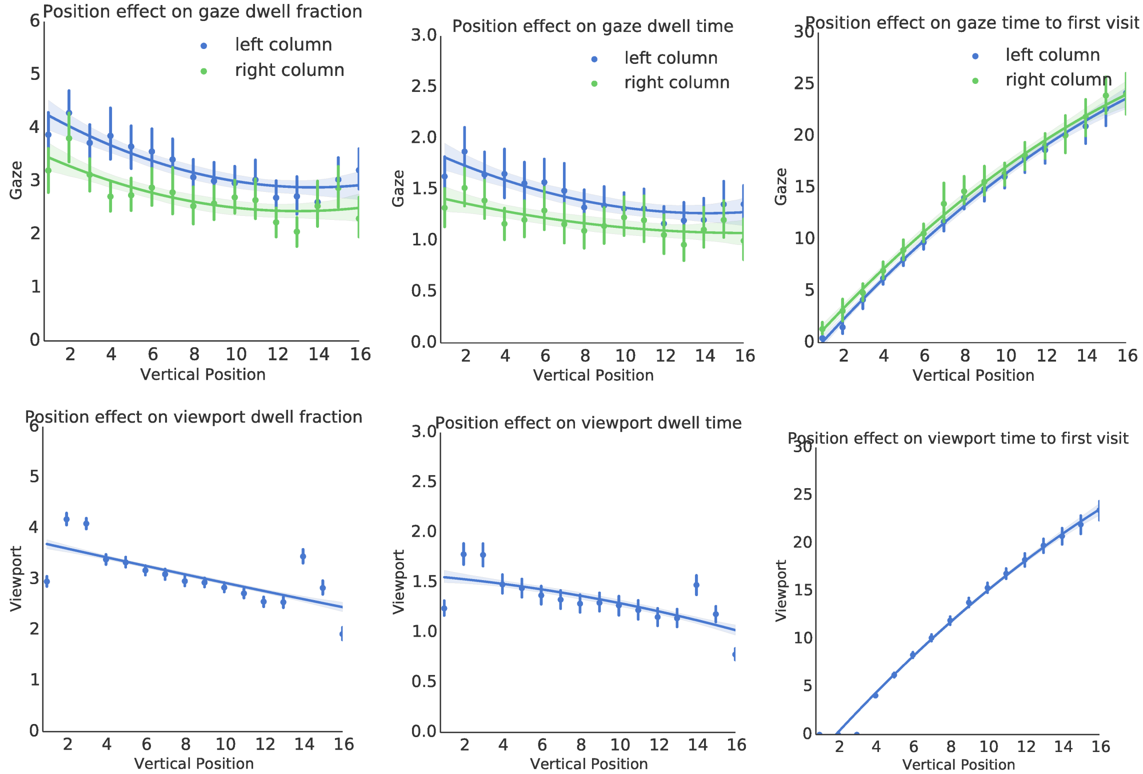


Figure 4: **The effect of vertical position on gaze metrics: dwell fraction % (left), dwell time in seconds (middle) and time to first in seconds visit (right). The blue and green lines show the metrics for the left and right column, respectively.**

panels in Figure 4, is that blue curves always stay above their green counterparts. This indicates that attention bias towards the left column is consistent, irrespective of the vertical position of content. In contrast, the left-column bias is no longer revealed when using viewport metrics – the blue and green curves overlap as seen in the bottom panels of Figure 4. The top-right panel in Figure 4 corresponds with the fact that most users first look at contents in the order from left to right, yielding blue curve below the green one.

We also observe the non-monotonic "boat-shape" curves in top-left and top-middle panels in Figure 4 – the gaze time and fraction first decrease w.r.t increasing position, and then slightly rebound towards reaching the bottom of the page. Similar trend can also be observed using viewport metrics (bottom panels of Figure 4), albeit suffering from a sharper growth towards the end. The most surprising observation is the bump at position 2, which corresponds to highest gaze time in seconds and in % amongst all positions. And such bump is consistent for both the left and right columns. One possible explanation for the bump at position 2 is the presence of short scrolls on mobile phones. Unlike desktop where the page up down keys allow users to move from one page fold to another non-overlapping page fold, in mobile phones, users often tend to perform short scrolls that may render the second or third result visible across multiple viewports and for longer time than the first result.

Furthermore, while the bump at position 2 is commonly observed using both gaze and viewport metrics, we find that the presence of bump at position 3 is unique to viewport (see bottom of Figure 4). We infer that this is again due to the drawback of viewport assigning equal likelihood for each position visible on screen, which does not reflect the actual biased attention distribution on the screen – on average almost 70% of the users' attention is focused on the top half of the phone screen, as previously reported in [16]. In other words, when both the second and third rows are visible on the phone screen, users' actual attention suffers from a decay (as reflected by the gaze metrics), instead of even split between the two rows (as reflected by the viewport metrics).

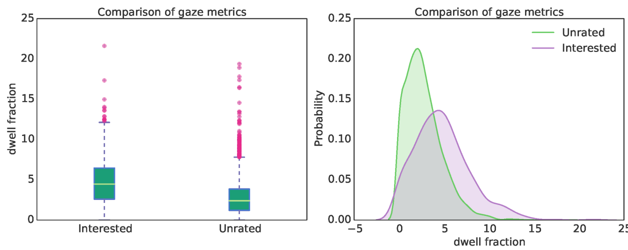
The findings of non-monotonic attention decay with rank position, as well as attention bias towards left column, may have implications for design of a novel personalization feed by prioritizing positions receiving significant portion of gaze attention.

6. GAZE REVEALS USER INTEREST

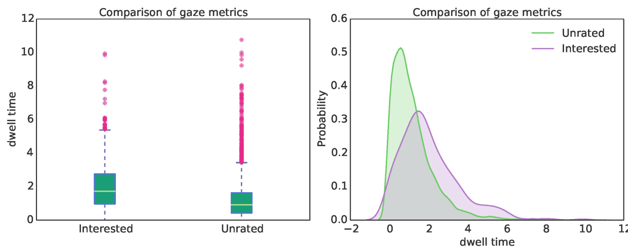
The rapidly growing population of smartphone users has created the need for better understanding and harnessing users' behavior beyond just clicks. A common challenge is to reliably identify users' interest in scenarios when click-stream data is sparse or unavailable. In this section, we validate the relationship between mobile users' implicit interest inferred from their eye gaze and explicit interest.

To start with, Figure 3 shows an example of heatmap generated based on a user's gaze, where the red rectangle overlay indicates the items that user selected after browsing[3]. It is quite intriguing to see the alignment between gaze and ground truth interest. To get statistically meaningful results, we further examine whether users tend to spend more

[3]Note that the red overlay is for visualization purpose only. Participants did not see the red rectangle during the study.



(a) Gaze dwell fraction



(b) Gaze dwell time

Figure 5: **Gaze metrics of dwell fraction (top) and dwell time (bottom) between unrated and interested items on Play Store game collection pages are statistically significantly different ($p < 10e{-7}$ in a two sided two-sample t-test) with highly non-overlapping 95% confidence intervals.**

attention on interested items, as opposed to unrated ones.

We split AOIs (across all collection pages and all participants) into two groups: the ones that a user exhibits interest in installing, and the remaining unrated ones. We perform a two-sample t-test on attention metrics between unrated AOIs and user interested AOIs. Figure 5 shows the estimated distribution of gaze metrics (dwell fraction and dwell time) between the unrated items (in green) and those items user indicates explicit interest (in purple). Participants spend on average $4.78 \pm 0.13\%$ gaze on items they are interested, as opposed to $2.76 \pm 0.03\%$ on unrated items. We observe substantial difference between gaze on interested and unrated items, as revealed by a t-test; $t(604, 4228) = 15.54$, $^{***}p < 0.001$.

Table 1 summarizes the results for all gaze and viewport metrics. For each metric, we report the mean and standard errors. The last two columns of the table shows the p-value and statistics using two sample t-test, allowing us to establish statistical differences in measured quantities among the metrics. All of the t-tests show significant differences ($p < 0.001$). It is also worth noting that the t-statistic derived from gaze signals is always higher than that of viewport metrics, suggesting gaze signals are more fine-grained in distinguishing interested vs. unrated items.

7. USER INTEREST INFERENCE MODEL

We have shown in previous section that gaze can be indicative signal of user self-reported interest in the context of digital media platform. The strong dichotomy of gaze signals between interested and unrated items leads us to a natural modeling and prediction question – can we reliably predict users' interest at AOI-level by employing multitude level of attention-based engagement metrics? In this section, we build upon previous insights and address this issue.

7.1 Methods

For digital collection pages such as in Google Play Store, users' interest in an individual item can be effected by multiple confounding factors: the position of the item, interestingness of the content itself, and users' intrinsic preference etc. In the past decade, web service providers have been using approaches such as collaborative filtering [25] to infer user interest by leveraging users' click log information or other types of engagement activities. However, mobile users nowadays spend significant portion of time purely browsing without clicking often – making prediction of users' interest in click-sparse scenario still a challenging problem. In contrast with widely adopted click based approaches, we adapt insights from previous sections and develop a probabilistic model that accounts for users' attention data including gaze and viewport. Specifically, we formulate this as a binary prediction problem: given the AOI of any item $\mathcal{I}_k^{(p,r)}$ indexed by r on a collection page p, a classifier predicts whether a user U_k is interested (positive) or not (negative). In our experiment, each AOI corresponds to a single game thumbnail. We consider the following comprehensive set of features that can be derived from our attention behavioral data.

- *Position*: The row-wise position of the item (a discrete variable varying from 1 to 16), and left vs. right column.
- *Gaze*: The gaze attention metrics introduced in Section 4.1, including TimeOnAOI (in seconds), %TimeOnAOI and TimeToFirstVisit.
- *Viewport*: The viewport-based attention metrics introduced in Section 4.2, including TimeOnAOI (in seconds), %TimeOnAOI and TimeToFirstVisit.
- *Categorical preference*: Users' rating of $C(\mathcal{I}_k^{(p,r)})$, where $C(\cdot)$ denotes the category of the item. In our cases, $C(\cdot)$ belongs to one of the following categories: word search, racing games, board games, kids painting.

Our initial training data consists of 604 positive examples and 4228 negative examples, aggregated across all participants and all 5 collection pages in use. The unbalanced size between positive and negative examples is due to that participant only marks *up to 5 items* as interested on each collection page with 32 items. We further conduct up-sampling and enlarge the positive example by 7 times, in order for matching the sample size of negative examples. We train binary classifiers using three methods: SVM with RBF kernel [11], Random Forest [19] as well as Decision Tree (with maximum depth 8). For each method, we perform 10-fold cross validation, and measure the accuracy by AUC (area under the ROC curve). Thus, we repeatedly train ten models and use them to obtain the predictions for each of the held-out set among the ten folds. We report the AUC computed from the ten test folds combined.

7.2 Experimental Results

Table 2 summarizes the AUC score of various inference models based on different set of features. There are several interesting observations can be drawn from the results. First, when using features from attention signals only, both gaze metrics and viewport metrics can perform better than

	Metrics	Unrated (mean ± std)	Interested (mean ± std)	p-value	T-test
Gaze	TimeOnAOI (seconds)	1.21 ± 0.02	2.05 ± 0.06	$^{***}p < 0.001$	12.93
	%TimeOnAOI	2.76 ± 0.03	4.78 ± 0.13	$^{***}p < 0.001$	15.554
	TimeToFirstVisit (seconds)	13.75 ± 0.17	12.24 ± 0.40	$^{***}p < 0.001$	-3.498
Viewport	TimeOnAOI (seconds)	1.30 ± 0.01	1.55 ± 0.04	$^{***}p < 0.001$	6.39
	%TimeOnAOI	3.00 ± 0.02	3.48 ± 0.04	$^{***}p < 0.001$	10.30
	TimeToFirstVisit (seconds)	12.11 ± 0.14	10.70 ± 0.37	$^{***}p < 0.001$	-3.55

Table 1: **Gaze and viewport metrics summarized for AOIs that users explicitly indicate interest vs. unrated (M±SE). Two sample t-test significance is annotated using the following coding: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.**

	Category	Position	Viewport		Gaze		Viewport & Gaze		All Features
			w/o Position	w/ Position	w/o Position	w/ Position	w/o Position	w/ Position	
SVM (RBF kernel)	66.36	54.57	60.50	66.38	69.02	75.56	75.54	82.90	**90.32**
Random Forest	65.76	57.82	69.90	75.44	77.26	78.60	79.40	80.62	84.83
Decision Tree	66.36	57.71	64.85	66.49	72.13	72.57	74.05	73.62	80.62

Table 2: **AUC (area under the ROC curve, %) of user interest inference model with various set of features. Note that a random guessing baseline would yield 50% accuracy on average.**

random guessing (69.90% and 77.26% respectively when using Random Forest classifier). In particular, using gaze attention signals alone can yield almost 8% higher AUC compared to using viewport signals.

Furthermore, we find that although positional feature by itself is not indicative enough to predict users' AOI-level interest (57.82%), it helps improve performance in general when combining with attention metrics. Interestingly, we notice that the relative improvement brought by adding positional feature is more significant for viewport than gaze. For example, when using Random Forest classifier, adding position feature to viewport features improves AUC by 5.5%, whereas adding it to gaze features improves AUC only by 1.34%. Relating to previous discovery in Section 5.1, we infer this is due to that viewport cannot differentiate between left and right column attention bias, thus adding positional feature can be beneficial in compensating for such limitation.

Besides using engagement related features (attention and position), we also find that having privileged information of users' categorical-level interest can be indicative signal of AOI-level interest. Using categorical rating itself can yield performance better than using viewport metrics (66.36% vs. 60.50% when using SVM classifier). When integrating all the features including categorical rating, our best classifier (SVM) gives AUC score as high as 90.32%. The result is very encouraging, given that it is based on purely attention signals without any click information.

8. FEED RANKING AND PERSONALIZATION WITH GAZE

Given the strong predictability of users' interest at AOI-level, we take one step further and investigate if we can personalize the ranking order of items and improve the feed relevance for a specific collection page, using attention related features.

8.1 Method

We model this as a bipartite ranking problem. To start with, we briefly review the basic SVM-ranking model [12]. Specifically, each training example corresponds to a page session $S = (S_+, S_-)$, where $S_+ = \{x_+^{(1)}, ..., x_+^{(m)}\} \in X^m$ are the positive examples corresponding to items a user is interested, and $S_- = \{x_-^{(1)}, ..., x_-^{(n)}\} \in X^n$ are the remaining items as negative examples. The goal is to learn a function $f \in \mathcal{F}$ which minimizes the following empirical loss

$$f^* = \min_{f \in \mathcal{F}} \Big(\sum_S (\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathcal{L}_{\text{hinge}}(f, x_+^{(i)}, x_-^{(j)})) + \lambda ||f||^2 \Big),$$

where

$$\mathcal{L}_{\text{hinge}} = \max \Big(1 - \big(f(x_+^{(i)}) - f(x_-^{(j)}) \big), 0 \Big).$$

Intuitively, an effective ranking algorithm assigns high ranking (relevance) score to those positive examples and low score to those negative ones, and push the difference between these two as far as possible. The ranking score can be used for re-ordering items and optimizing feed for a given page.

For each game thumbnail AOI $x^{(i)}$, we extract attention features the same way as used for the user interest inference model. After removing non-usable data, our data collection consists of 144 page browsing sessions in total, aggregated all participants and all 5 collection pages each. We randomly hold out 36 pages for testing, and train the ranking model on the remaining 108 page sessions containing in total 14,261 pairwise ranking preference $(x_+^{(i)}, x_-^{(j)})$.

8.2 Experimental Results

We adopt NDCG (Normalized Discounted Cumulative Gain) metric [10] for evaluating the ranking models. For a page browsing session, we evaluate the reranking of game collection page based on users' explicit interest feedback. NDCG

	Position	Viewport w/o Position	Viewport w/ Position	Gaze w/o Position	Gaze w/ Position	Viewport & Gaze w/o Position	Viewport & Gaze w/ Position	All Features
NDCG@3	0.176	0.257	0.322	0.413	0.444	0.465	0.452	**0.594**
NDCG@5	0.184	0.232	0.281	0.405	0.410	0.410	0.406	**0.513**
NDCG@10	0.322	0.414	0.434	0.538	0.576	0.567	0.577	**0.695**

Table 3: **Ranking performance using multitude attention based features.**

metric varies from 0 to 1, with 1 representing the ideal ranking of the items. This metric is commonly used in information retrieval literature for evaluating the performance of information retrieval systems [20]. To compute the NDCG@k we use logarithmic position discount:

$$NDCG@k = \frac{1}{IDCG}\sum_{i=1}^{k}\frac{2^{rel_i}-1}{\log_2(i+1)},$$

where k denotes the number of entities that can be recommended, and rel_i being the ranking relevance score of item at position i produced by the algorithm.

The ranking performance of $NDCG@k$ for $k = \{3, 5, 10\}$ is reported in Table 3. We find that using gaze metrics can be substantially more advantageous than viewport metrics. In the cases when k is small (e.g, $k = 3, 5$), ranking with gaze outperforms viewport by a large margin > 0.15. This reassures that gaze-based attention metrics can be more effective not only in task for AOI-level interest inference, but also in page-level retrieval tasks.

Similar to what we observed in previous section, here we also note that the relative improvement gained by adding positional feature is more considerable for viewport than gaze. For example, adding position information on viewport features can boost $NDCG@3$ from 0.257 to 0.322; whereas for gaze features, the gain is comparably small (0.031). Again, we believe this might be the inadequacy of viewport discerning the position bias in multi-column layout pages seen in digital media sites.

When combining all the features in hand, our attention-based ranking model attains $NDCG@10$ of 0.695, which means the collection page has been personalized in a way that effectively pushes items attracting to the user on top.

To the best of our knowledge, this is the first study demonstrating the efficacy of employing attention engagement metrics for ranking task in the context of digital media platform. We envision that incorporating attention signals can be complementary to many other click-based ranking algorithms. And we plan to investigate this as part of future work.

9. DISCUSSION

Our preliminary findings raise many important open questions that would be interesting to take into account in future research. First, it would be interesting to run online experiments, optimizing the collection feed ranking in real time, and evaluate the efficacy of tuning the order of displayed items iteratively with attention features. Second, our user interest inference model can be extended to combine with contextual information such as image saliency of the thumbnails [9]. Saliency-based model has been recently applied to analyze and predict user examination pattern on SERP [21]. We plan to investigate how such saliency based features can be incorporated into current model, and further improve the prediction accuracy. Third, currently the interest inference model is purely based on generic attention features, and might be potentially transferable to domains such as movie and music collection pages etc. Future work involves conducting transferability test on other domains.

Admittedly, our experiments are simulated in a rather simplified laboratory environment, which might not entirely reflect the realistic browsing situations. In order to make the setup more applicable, future lab study can be designed to allow multi-session browsing, where participants can search and click on the icons freely (which will direct users to the app detail page). This can lead to further insights on how can we integrate click signals with attention signals in order for better predicting and modeling users browsing behavior on mobile platforms.

10. CONCLUSION

In this paper, we have presented the first quantitative eye-tracking study analyzing the relationship between mobile user's implicit eye gaze and explicit interest in novel contexts such as digital collection pages on Google Play Store. We find significantly different distribution of gaze metrics on items that a user rate as interesting vs. not (e.g., longer gaze time on interesting items vs. unrated ones). Built upon this insight, our purely attention based interest inference model is able to attain AUC score as high as 90.32% in predicting user's interest in individual items, in digital collections pages consisting of 30+ items. In addition, we also show the promise of improving feed relevance and personalizing the order of displayed items on collection page, using various attention related features. These findings have implications for the design of a novel personalization and recommendation mechanism by (1) prioritizing items that are most likely of interest to the user based on historical attention behavior, and (2) prioritizing positions receiving significant portion of gaze attention.

11. ACKNOWLEDGMENTS

We thank Julie Fischer and Khalil Cader for helping with preparing study materials and formalizing the research questions. Yixuan Li has been supported by US Army Research Office W911NF-14-1-0477 for her thesis research.

12. REFERENCES

[1] D. Bosomworth. Mobile marketing statistics 2015. *Smart Insights site*, 2013.

[2] H. Bota. Nonlinear composite search results. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 345–347. ACM, 2016.

[3] G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to

predict salient regions of web pages. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 21–30. ACM, 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] G. Buscher, L. van Elst, and A. Dengel. Segment-level display time as implicit feedback: a comparison to eye tracking. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 67–74. ACM, 2009.

[6] M. Claypool, P. Le, M. Wased, and D. Brown. Implicit interest indicators. In *Proceedings of the 6th international conference on Intelligent user interfaces*, pages 33–40. ACM, 2001.

[7] Z. Guan and E. Cutrell. An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420. ACM, 2007.

[8] J. Huang and A. Diriye. Web user interaction mining from touch-enabled mobile devices. In *HCIR workshop*. Citeseer, 2012.

[9] L. Itti, C. Koch, E. Niebur, et al. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[10] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[11] T. Joachims. Making large scale svm learning practical. Technical report, Universität Dortmund, 1999.

[12] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the 8th ACM SIGKDD conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.

[13] J. Kim, P. Thomas, R. Sankaranarayana, T. Gedeon, and H.-J. Yoon. Eye-tracking analysis of user behavior and performance in web search on large and small screens. *Journal of the Association for Information Science and Technology*, 66(3):526–544, 2015.

[14] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba. Eye tracking for everyone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2176–2184, 2016.

[15] D. Lagun and E. Agichtein. Viewser: Enabling large-scale remote user studies of web search examination and interaction. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 365–374. ACM, 2011.

[16] D. Lagun, C.-H. Hsieh, D. Webster, and V. Navalpakkam. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 113–122. ACM, 2014.

[17] D. Lagun and M. Lalmas. Understanding user attention and engagement in online news reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 113–122. ACM, 2016.

[18] D. Lagun, D. McMahon, , and V. Navalpakkam. Understanding mobile searcher attention with rich ad formats. In *The 25th 2016 ACM Conference on Information and Knowledge Management*. ACM, 2016.

[19] A. Liaw and M. Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[20] T.-Y. Liu, J. Xu, T. Qin, W. Xiong, and H. Li. Letor: Benchmark dataset for research on learning to rank for information retrieval. In *Proceedings of SIGIR 2007 workshop on learning to rank for information retrieval*, pages 3–10, 2007.

[21] Y. Liu, Z. Liu, K. Zhou, M. Wang, H. Luan, C. Wang, M. Zhang, and S. Ma. Predicting search user examination with visual saliency. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 619–628. ACM, 2016.

[22] 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. ACM, 2015.

[23] 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. ACM, 2013.

[24] Q.-X. Qu, L. Zhang, W.-Y. Chao, and V. Duffy. User experience design based on eye-tracking technology: A case study on smart phone apps. *VG Duffy, Applied Digital Human Modeling and Simulation, Springer*, 2016.

[25] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[26] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao. Turkergaze: crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755*, 2015.

[27] L. Yang, O. Toubia, and M. G. De Jong. A bounded rationality model of information search and choice in preference measurement. *Journal of Marketing Research*, 52(2):166–183, 2015.

[28] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling. Appearance-based gaze estimation in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4511–4520, 2015.

[29] Q. Zhao, S. Chang, F. M. Harper, and J. A. Konstan. Gaze prediction for recommender systems. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 131–138. ACM, 2016.