Friendly, Appealing or Both? Characterising User Experience in Sponsored Search Landing Pages

Marc Bron
Yahoo!

Miriam Redi
Bell Labs

Mounia Lalmas
Yahoo!

Fabrizio Silvestri
Facebook

Huw Evans
Yahoo!

Mahlon Chute
Yahoo!

ABSTRACT

Many of today's websites have recognised the importance of mobile friendly pages to keep users engaged and to provide a satisfying user experience. However, next to the experience provided by the sites themselves, advertisements, when clicked, present users with landing pages that are not necessarily mobile friendly. We explore what type of features are able to characterise the mobile friendliness of sponsored search ad landing pages. To have a complete understanding of the mobile ad experience in terms of layout and visual appearance, we also explore the notion of the ad page aesthetic appeal. We design and collect annotations for both dimensions on a large set of ads, and find that mobile friendliness and aesthetics represent different notions. We perform a comprehensive study of the effectiveness of over 120 features on the tasks of friendliness and aesthetics prediction. We find that next to general page size, HTML, and resource usage based features, several features based on the visual composition of landing pages are important to determine mobile friendliness and aesthetics. We demonstrate the additional benefit of these various types of features by comparing against the mobile friendliness guidelines provided by W3C. Finally, we use our models to determine the state of landing page mobile friendliness and aesthetics on a large sample of advertisements of a major internet company.

1. INTRODUCTION

Many free web services generate revenue by presenting users with advertisements (ads for short) in addition to the services they provide. In this work, we focus on search engines, where users are often served ads, together with organic search results as answer to queries submitted to the search engine. This is referred to as *sponsored search*, and has been extensively studied for many years in the context of desktop search [3, 33, 34]. When shown results in response to his or her query – usually a mixture of organic and sponsored results – a user may decide to click on the latter, i.e., the ad. After an ad click, the user is redirected to the ad *landing page*, which is either a web page specifically created for that ad, or the advertiser homepage [4].

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054193

Previous research on sponsored search has mostly focused on predicting how an ad will perform according to various effectiveness metrics, specifically click-through rate [3, 4, 6, 15]. In our work, we look at how users *experience* the ad landing pages in the context of sponsored search. It is well known that the way users experience an ad landing page, i.e., the *ad post-click experience*, is an important factor of the quality of the ad. Indeed, a negative post-click experience can have disruptive consequences on the overall number of visitors and therefore on total revenue [11].

For our study, we specifically focus on the context of advertising in *mobile* devices. The limited screen and resources of mobile phones have created new challenges for advertising, thus completely re-designing the way in which users consume the ad. There have been various efforts looking at the post-click experience in (mobile) advertising, finding, for example, that dwell time is a good proxy of an ad post-click experience [19]. However, recent work on dwell time estimation [1] found that historical dwell time performance of ads is crucial for accurate dwell time prediction. Since advertisers continually adjust their ads (and landing pages) a large portion of ads is cold (does not have historical information about dwell time, clicks, etc.). Further, there are various interpretations of dwell time, e.g., relevance, quality, interest. Without knowing the reason for ad pages' low dwell times, remedying its lack of performance is difficult.

Our paper therefore aims to characterise two specific aspects of the post-click experience: (i) the *mobile friendliness* of ad landing pages and (ii) their *aesthetic appeal*.

A general web page is mobile friendly if it has a *good user experience on a mobile device*, where a good experience is a combination of great performance and mobile specific experience. Since we are interested in advertising, we re-shape the notion of mobile-friendliness and adapt it for *ad* landing pages. We define an ad landing page as *ad mobile friendly* if (1) it provides a good interactive experience (e.g. big buttons, few links), (2) it makes it easy to understand what the ad is about (e.g. the product advertised) and (3) allows the user to convert (e.g. purchase the product advertised) *when* shown on a mobile device.

We further hypothesise that beautiful ad landing pages provide a better user experience than pages that do not have this characteristic and investigate a previously unexplored dimension of a mobile ad landing page: its *aesthetic appeal*. As defined in [18], aesthetic appeal is concerned with the sensory and visual appeal of an interface and is seen as an important factor for engagement. Aesthetic appeal has been shown to manifest in a site screen layout, graphics, and use of design principles. It has also been suggested that aesthetics promote focused attention and stimulate curiosity, two important components of user engagement.

Previous work [21] simply characterized sponsored search ad landing pages in terms of recommendations by the W3C consortium, which are designed for general web pages. In this work we carry out for the first time a detailed, large scale study of ad landing page mobile friendliness and aesthetic appeal, and analyse how these two dimensions relate. We contribute to the field in the following ways.

- We design a set of novel editorial guidelines for mobile friendliness and aesthetics that account for the fact that we are interested in the advertising domain. We collect around 4,000 annotations, and find that both mobile friendliness and aesthetics, although related, represent different dimensions of the ad mobile page experience.
- We explore a set of more than 120 features able to characterise mobile friendliness and aesthetics of ad landing pages. We start out with a comprehensive set of features based on the landing page itself, such as its length, its structure, and viewport information, recently used in dwell time prediction work [1]. We also explore additional features specifically designed to detect mobile friendliness as well as visual features inspired by work in computational aesthetics [23, 39].
- We then explore several predictive models and feature set combinations to gain an understanding of the features and to show their significance in predicting ad mobile friendliness and aesthetic appeal as compared to the W3C recommendations designed for generic web pages.
- We find that both visual and intrinsic page features are important to predict mobile friendliness and aesthetics. We show that HTML and W3C features that capture the structure, the size of the page and amount of style sheets are good at predicting mobile friendliness, but bring little in predicting aesthetic appeal. From a visual feature perspective, we found a big impact of color brightness on user experience, and that, unlike artistic pictures, the uniqueness of the ad landing page composition negatively impact both aesthetic appeal and mobile friendliness.
- Finally, we found that on a large sample of ads, mobile friendly pages are positively associated with long clicks whereas aesthetically appealing pages are positively associated with conversions.

2. RELATED WORK

Our work belongs to the field of advertising and computational aesthetics. We position our work within these areas.

Computational aesthetics. Computational aesthetics is a branch of computer vision that develops frameworks able to automatically score images in terms of beauty and designs features inspired by photographic rules and describable stylistic attributes [8, 14, 16, 25]. Similar to our work, such features have been used to infer other abstract dimensions related to the subjective perception of the visual world. Examples include image affective value [23], video creativity [29] and interestingness [14]. Similar tools have also been used to predict image memorability [13] and popularity [17]. Different from previous work e.g. [30, 32, 38], which applies to photographic material, we use aesthetic features over landing page screen-shots in the context of mobile advertising.

Low-level visual features have been used to evaluate affective responses to web pages [2, 31, 37]. Regarding the perception of web pages, aesthetics have been shown to be particularly significant for user satisfaction and pleasure [20]. Also, aesthetics was shown to compensate for usability and vice versa, depending on the context [9]. In non-serious contexts, aesthetics can improve user perception, whereas in serious tasks, improving usability can improve the perception of the information quality. In our work, instead of looking at general web pages, we focus on the domain of landing pages used in mobile advertising. In addition, while previous work evaluates how user experience is influenced by a particular type of low-level features, we experiment with a wide range of feature types, visual *and* HTLM features, and evaluate their respective merit for mobile friendliness and aesthetic prediction of ad landing pages.

In the context of computational advertising, visual features have been used before to predict ad click-through rate (CTR) [5, 24] and detect offensive ads [39]. Our work differs in that we use visual features to predict new dimensions of online advertising, the mobile friendliness and aesthetic of ad landing pages.

User Experience and ad landing pages. CTR – the number of times the ad was clicked out of the number of times it has been shown – is a common measure of an ad "performance". However, CTR does not account for how users experience the ad when they land on the ad site, namely their *post-click experience*. A positive experience increases the probability of users "converting" (e.g., purchasing an item, or registering to a mailing list). A positive post-click experience does not necessarily mean a conversion, as there may be many reasons why a conversion does not happen, independent of the quality of the ad served to users. In addition, conversion rates have been shown to differ significantly depending on the type of landing page [3], as well as that conversion rates are generally quite low making it difficult to obtain reliable estimates in small to moderate size samples.

A good proxy of the post-click experience is the time a user spends on the ad site before returning back to the publisher site, where "the longer the time, the more likely the experience was positive". Dwell time has been shown to be a good proxy of an ad post-click experience [19]. In our work, we also show the relationship between dwell time and mobile friendliness and aesthetic appeal. Finally, in the context of mobile advertising, whether the landing page of an ad is mobile-optimised or not was shown to affect the post-click experience [21]. Our research adds to this body of work by analysing features of landing pages focusing on characterising the mobile friendliness and aesthetics of the landing pages.

Mobile Friendliness. Online tools with a similar aim as our work use various techniques to score landing pages (given their URLs) in terms of mobile friendliness. Major companies such as Google and Bing[1] provide online frameworks that test whether pages are mobile friendly, including feedbacks and suggestions when they fail the test. Similarly, the W3C consortium has built its own *mobileOK* checker.[2] Although inspired by these tools, our work is fundamentally different. First, the commercial tools do not provide transparent studies regarding what makes web pages mobile friendly, even though these tools come with design guidelines for mobile friendly pages. Moreover, unlike our work, these tools are not specifically tailored for advertising. In this paper, we design an end-to-end system that models the specific notion of *ad mobile-friendliness*, a notion embracing both the user experience and the commercial dimension of the landing pages. Also, the existing tools discussed above are based on text or HTML features only. In our framework we explore the visual aspect of mobile friendly ad landing pages by using computational aesthetic features. We also look at the aesthetic dimension of an ad landing page.

To demonstrate the added value of the various types of features used in our framework, we compare it with a baseline built from the

[1] `https://www.google.com/webmasters/tools/mobile-friendly/`, `https://www.bing.com/webmaster/tools/mobile-friendliness`
[2] `https://validator.w3.org/mobile/`

output of the W3C MobileOK tool, one of the few available mobile friendliness checker allowing for batch processing.

Other work has explored the notion of mobile friendliness outside the context of advertising. Liu et al. [21] compare desktop and mobile landing pages using the mobileOK tool. Also, Qian et al. [28] analyze the resources needed for mobile web browsing in terms of bandwidth and energy. While these methods use existing tools to analyze the mobile friendliness of general web pages, we design a framework with classifiers and features specifically tailored to mobile advertising.

3. METHOD

We describe the features and models used to determine the mobile friendliness and aesthetic appeal of mobile sponsored search ad landing pages. We focus on two types of features: (i) based on the markup elements and text in the page (HTML features); and (ii) derived from an image representation of the landing page (visual features). To obtain pages as they would be seen by users we render them using a mobile browser before scraping the pages.

3.1 HTML based features

One source of features is the markup, text, and objects in the rendered HTML of the pages. Table 1 shows the features we extract divided into five categories, following [1, 19].

Mobile optimized: captures whether a page is specifically designed for mobile or whether it is a desktop page. We included features suggested in [19] to classify mobile optimized pages, such as whether the page contains mobile specific elements. We also added features that capture page size in terms of text, media, and style.

Window size: capture only aspects of the size of the rendered HTLM image and may be used to detect whether the size of a page is suitable for mobile devices.

Readability: used to identify the formality of the language used in the landing page text. The intuition is that dense and formal texts may be less pleasant to read on a mobile device. These features were found to be significant in predicting dwell time [1].

Input: pages that require users to provide information through many forms may be considered less mobile friendly than pages that require less input. These features capture the number and type of input elements in a page.

Navigation: captures the proportion of internal links, external links, and text contained in a page; e.g. mobile friendly pages may provide access to different sections of a page through internal links.

Feature	Description
Mobile Optimized	
LandingTextLength	No. non HTML element characters in the page.
LandingMainTextLength	No. non HTML element characters in the page with boilerplate text removed.
LandingTextMainTextRatio	No. characters with and without boilerplate text ratio.
HtmlClickToCallAttribute	Is there a click to call button?
HtmlIphoneButtonAttribute	Is there an iPhone button?
HtmlMetaViewportExisted	Is viewport available?
HtmlNumImages	No. of images contained in landing page.
HtmlMediaAttribute	Is there a media (e.g., video) on landing page?
CSS_COUNT	No. of CSS style sheets loaded.
FRAME_COUNT	No. of frames in the page.
Window Size	
MAIN_ORG._HTML_SIZE	No. characters in the page including HTML elements.
HtmlWindowSize	Size of the window.
IMAGE_WIDTH	Width of the rendered landing page.
IMAGE_HEIGHT	Height of the rendered landing page.
Readability	
NumWordsHtml	No. of words in the page.
NumSyllablesHtml	No. of syllables in the page.
NumComplexHtml	No. of complex words (3 or more syllables) in page.
FogReadabilityHtmlScore	Gunner fog index
FleschReadabilityHtmlScore	Flesh score
KincaidReadabilityHtmlScore	Flesh-Kinkaid score
Inputform	
HtmlNumClickable	No. of clickable objects int he landing page.
HtmlNumDropdown	No. of dropdown elments.
HtmlInputRadioCount	No. of radio buttons.
HtmlInputTextCount	No. of Input Strings.
HtmlInputCheckboxCount	No. of checkbox.
HtmlOnclickCount	No. of java script triggers.
Navigation	
LinksExternalCount	No. of links pointing to external domains.
LinksInternalCount	No. of links pointing to same domain as landing page.
LINKS_COUNT	Sum of the previous two features.
LinksExternalInternalRatio	Ratio of External vs. Internal links.
LinksExternalTotalRatio	fraction of external links.
LinksInternalTotalRatio	fraction of internal link.
LinksTextLengthExternalRatio	Text per external links ratio.
LinksTextLengthInternalRatio	Text per internal links ratio.
LinksTextLengthTotalRatio	Text per total number of links ratio
LinksMainLengthTotalRatio	Main Text (no boilerplate) per total no. of links ratio.
LinksMainLengthExternalRatio	Main Text (no boilerplate) per External links ratio.
LinksMainLengthInternalRatio	Main Text (no boilerplate) per Internal links ratio.

Table 1: Features extracted from the rendered HTML of sponsored search ad landing pages.

3.2 Image features

Aesthetics can play an important role in user engagement [18]. To understand visual aesthetic attributes of ad landing pages, we adopted a set of visual features from the computational aesthetics field, shown in Table 2. Specifically, we extract the following features from the screenshot of the rendered ad page:

Colors The color palette used in a mobile page affects its aesthetic perception and its credibility [27]. To compute the color distribution, we convert the screenshot to the *Hue, Saturation, Brightness (H,S,V)* color space, and compute the *Average* of each channel (H,S,V) for the whole image and for the image *Central Quadrant*. We then compute the *HSV Color Histograms* and the *HSV Contrasts* [23], by quantizing the Hue channel into 12 bins, the Brightness channels into 5 bins, and the Saturation channel into 3 bins.

Textures and Contrast Textural patterns in the page can influence affective reactions to web pages [37]. We characterize page texture by computing GLCM properties such as *Entropy, Energy, Homogeneity, Contrast* using Haralick's features [10]. We also compute a *Contrast* metric that quantifies the extent to which object contours are distinguishable.

Image Quality Registration, manipulation, and encoding can degrade the overall image quality. To quantify the level of image integrity after processing, we extract a set of image quality metrics that have been found useful to evaluate the quality of native ads [39]. We compute the *Contrast Balance*, the *Exposure Balance*, a metric evaluating the quality of the image after JPEG compression (*JPEG Quality* [36]) and a *JPEG Blockiness* feature detecting the presence of blocking artifacts. Further, *Pleasure*, *Arousal*, and *Dominance* are three features that correspond to emotional coordinates based on brightness and saturation [23]. Finally we compute the overall image *Sharpness* [39] on the whole image and on the *Foreground* only.

Image Layout Object distribution and symmetry play an important role in interface design and usability [2]. We look at the overall image layout and extract a set of compositional features to understand the object arrangement in the page space. From previous

Feature	Description
Colors	
H,S,V [23]	Average *Hue, Saturation, Brightness* computed on the *whole* image
H,S,V [23] (Central Quadrant)	Average *Hue, Saturation, Brightness* computed on the the *central quadrant*
H,S,V Color Histograms [23]	Histograms of H, S and V values quantized over 12, 3, and 5 bins
H,S,V Contrasts [23]	Standard deviation of the HSV Color Histograms distributions
Textures	
Contrast [39]	Ratio between the sum of max and min luminance values and the average luminance
GLCM Properties [23]	Entropy, Energy, Contrast, and Homogeneity of the Gray-Level Co-Occurrence Matrix
Quality	
Contrast Balance	Distance between original and contrast-normalized images
Exposure Balance	Absolute value of the luminance histogram skewness
JPEG Quality	No-reference quality estimation algorithm in [36]
Sharpness	Sum of the image pixels after applying horizontal/vertical Sobel masks
Foreground Sharpness	Sum of the image pixels after applying horizontal/vertical Sobel masks
Pleasure, Arousal, Dominance	3 approx emotional coordinates based on brightness and saturation [23]
Layout	
Presence of Objects [39]	Amount of saliency [12] in 9 image quadrants
Uniqueness [39]	Difference between the image spectral signal and the average spectrum of natural images
Symmetry [39]	Difference between the HOG [7] feature vectors of the image left-half and right-half

Table 2: Visual features extracted from the screenshots of the ad landing pages.

computational aesthetic work [39], we compute a metric quantifying the *Presence of Objects* in 9 image areas, the *Symmetry* in the image, and a *Uniqueness* metric reflecting the unconventionality of the image composition.

3.3 Models

Our goal is not to compare the effectiveness of different machine learning methods or to develop new ones, but to find models for mobile friendliness and aesthetics that best fit the data in our current setting and that generalize well beyond our current sample. To this end we use four different machine learning methods with complementary properties. The eventual mobile friendliness and aesthetics models will allow us to characterise the current state of the mobile friendliness and aesthetics of mobile ad landing pages in the inventory of a large internet company.

The first model is **Multiclass logisitic regression**, which models the response variable Y by estimating the parameter $p_k(a)$ of a Bernoulli distribution using the inverse logit of a linear combination of the input features: $P(Y = 1|\vec{x}(a)) \sim Ber(p(a))$, with $p(a) = e^{\vec{b}\cdot\vec{x}(a)}/1 + e^{\vec{b}\cdot\vec{x}(a)}$. For the multiclass case we apply a one-versus-all strategy. We estimate $p_k(a)$ of the distribution modeling the response, where k is one of the possible class labels. Each random variable Y_k has a response of 1 when class label k is observed and 0 otherwise. We then pick the class for which $p_k(a)$ is maximum: $k = \mathrm{argmax}_{k\in K}\, p_k(a)$ where K is the set of classes.

The second model is **Gradient Boosted Machines (GBDT)**, an ensemble based classification method that aimed at discovering non-linear features from the data by sequentially fitting the error of the logistic loss function. $p_k(a)$ is estimated by sequentially fitting M decision trees for each class k: $p_k(a) = e^{F_k(a)}/\sum_{i=1}^{K} e^{F_i(a)}$ where $k = 1, K$ and $F_k = \sum_{i=1}^{m} wd_i(a)$ where $d(a)$ is a decision tree fit to the negative gradient of the loss function at each step i and w is a weight on the model contribution. We also use **Gradient Boosted Regression Trees (GBRT)**, a variant of GBDT that estimates the conditional mean (corresponding to a class label) of each segment of data, thus not requiring a separate model to be learned for each class, and reflecting the ordinal nature of the classes.

The fourth model is **Random Forest**, which is composed of an ensemble of independent decision trees fully grown on bootstrapped samples, one for each tree, of the training set. Each tree outputs a possible class and the final classification is given by the outcome of the majority of the trees in the ensemble.

4. DATA COLLECTION

To collect ground truth for ad mobile friendliness and aesthetics prediction, we designed a set of guidelines specifically tailored for our task. We collected professional editorial judgments for $4,000$ mobile ads sampled from a monthly log of ads impressed on mobile sites of Yahoo, a large Internet company.

4.1 Guidelines

Various guidelines to design mobile friendly websites are available.[3] These guidelines aim to convey best-practices and recommendations for app and web designers, i.e., particular design patterns, functionality, and features that should or should not be used. Such guidelines, however, do not specify how to score a particular landing page based on these recommendations. In this work, we aim to design a system able to score ads in terms of their mobile friendliness and aesthetic appeal. We thus need to collect *judgments* for a sample of ad landing pages representative of various page implementation styles. We therefore designed guidelines that help judges evaluate the mobile friendliness and aesthetic appeal of online ads. Although inspired by existing recommendations for mobile web designers, our guidelines are explicitly designed to address issues relevant to advertising: mobile friendly pages provide a good experience on mobile devices, whereas mobile friendly *ads* provide a good experience on mobile devices *and* allow users to easily consume the product advertised. To capture these variations, including aesthetic appeal, we use a graded schema.

Mobile Friendliness Grades. After a set of pilot experiments, we chose a 4-point scale to evaluate ad mobile friendliness:

- **Bad:** These pages provide a mobile experience of extremely low quality. In practice, these are often desktop pages "squeezed" on a mobile device screen.
- **Fair:** Pages that satisfy only 2 of the 3 requirements for good mobile friendly pages (see below).
- **Good:** Pages that satisfy the following criteria: Good readability of the content (e.g. uncluttered, big text; proportioned media size); High usability on a mobile device (e.g. big buttons, few links); and Simplified navigation (e.g. simple menus, clickable back or home buttons, visible call-to-action)
- **Perfect:** Mobile pages that satisfy all criteria for *good* mobile friendly pages, and in addition, are good *ads* on mobile. They provide ease of conversion (e.g. few-step conversion process, easy-to-fill form), and a good *product experience*, clearly explaining the product advertised and facilitating the user-business interaction.

The first three grades follow general *mobile friendliness* criteria, i.e., how easy it is to explore the page on a mobile device. The final

[3]See, for example, guidelines from W3C https://www.w3.org/2007/02/mwbp_flip_cards and Google https://developers.google.com/webmasters/mobile-sites/#why

grade (perfect) focuses on the "advertising" aspect of the mobile page and is concerned with whether the page facilitates conversion, e.g., purchase of the advertised product.

Aesthetically Pleasing Annotation. We also collected judgments on the *aesthetic appeal* of ad landing pages. Annotators were asked to express a binary judgement about the visual appeal of the landing page, based on their personal taste regarding the overall layout, color balance and composition.

4.2 Annotations

Annotation task. Judges were presented with a landing page URL as well as the four point mobile friendliness and the binary aesthetics scales. Judges clicked the URL which rendered the landing page in a mobile device emulator allowing full interaction with the page. In many hardware and software combinations available in the mobile ecosystem prevents comprehensive judging of pages on physical devices. The added advantage of an emulator is that it makes the judgement process more reproducible, as it is easier to restart a simulator than to obtain a particular physical device [35].

Judges interacted with the page –and if necessary clicked through to subsequent pages– to determine the mobile friendliness and ease of conversion of the ad landing page. When ready, the judges selected a radio button corresponding to their judgment of the page mobile friendliness. Judges were provided with an extensive guideline document specifying good and bad examples as well as borderline cases of the mobile friendliness criteria. Similar guideline documents have been published by W3C and Google. Using these guidelines, pages were assigned the appropriate rating based on the number of criteria they satisfied and the ease of conversion of the page. Further, if a judge decided that the site met the guideline requirements for aesthetic appeal, the "Is this mobile page aesthetically pleasing?" checkbox was selected.

Judgement procedure. During an initial pilot test, each judge reviewed one landing page. We observed a large variance in distribution of mobile friendliness ratings between judges. This caused concern that some judges might be applying guidelines more or less strictly than intended. In the second pilot test, 4 judges judged each 100 landing pages, and we ended with 66% of the landing pages having at least one disagreement, 39% with exactly one disagreement, and 27% with more than one disagreement. Based on the two pilots, we decided on the following procedure to increase consistency and accuracy of judgments. Two judges assessed each landing page as described above, then disagreements on the "mobile friendliness" judgment were resolved by a third judge, who could choose one of the original judgments or assign a new final judgment. The arbiters judges were a subset of the original group of judges, who had more experience with testing and whose judgment distribution was closer to the mean. This reduced annotation costs versus requiring three initial judgments. Through this procedure we obtained high agreement on the mobile friendliness judgements. The judgement of "aesthetically pleasing" is of a more subjective nature. Therefore if any judge marked a page as "aesthetically pleasing" it was labeled as such.

Judges. Judges were employees at Yahoo, the Internet company where this research was carried out. They were members of a team trained to accurately follow test guidelines and experienced with the testing environment and other tests of content quality. In total 17 individual judges were involved.

Statistics on judgement procedure. Each landing page judgment involved looking at and clicking around the landing page, assigning a mobile friendliness rating, deciding whether the page was aesthetically pleasing, and writing notes if required. This took around two minutes on average. In total, 1,633 landing pages were judged differently by the two initial judges and so required arbitration by a third judge. This gave us 63% agreement in our initial mobile friendliness judgments. Of the 37% disagreements, 30% were by one rating level, 5% were two or more rating levels, and 2% involved a Not Judged rating. We considered one-level disagreements expected due to the nature of the test. Two-level disagreements may indicate judge error and were given additional attention in arbitration. Disagreements involving NJ ratings were generally the result of the landing page being unavailable (404 error).

Descriptive statistics. We collected judgments for 4,025 pages. A majority of them were annotated as being *fair* or *good*. Less that 10% of the ads are perfect mobile friendly pages, 35% and 40% *fair* and *good* respectively, while a substantial number of ads provide a very *bad* mobile experience (16%). Finally, 5% of the ads were annotated as *aesthetically pleasing* by at least one of the 17 judges.

We find that the correlation between the mobile friendliness and the aesthetics judgements $\rho = 0.27$ using Spearman's ρ is significant $p < 0.001$, suggesting that aesthetic appeal brings another dimension to the quality of an ad landing page. Although most of (48% and 47%, respectively) of the aesthetically pleasing pages fall in the perfect and good mobile friendly classes, not all perfect mobile friendly pages are visually appealing.

We further find a positive correlation between mobile friendliness and dwell time (0.26), while the correlation between aesthetics and dwell time is 0.16. Dwell time is often used as a proxy of user engagement with ad landing pages. This suggests that users tend to stay longer on both mobile friendly and aesthetics pages. Although the correlation is statistically significant, it remains weak to moderate, indicating that mobile friendliness does not equate to dwell time and is adding a different dimension to the characterization of ad post-click experience. We did not include statistics for conversions as the sample is too small to obtain robust estimates.

5. RESULTS AND ANALYSIS

5.1 Experimental Setup

Train-test splitting. In sponsored search, advertisers bid on a set of keywords in an auction for ad impression slots. Each slot is associated with a particular keyword and if an advertiser wins a slot for any of the keywords then its ad is shown. The landing page associated with the ad may be different depending on the keyword for which the ad was shown, e.g., a beauty product company may show a landing page with a sunscreen product for the keyword "holiday" and a night cream product for the keyword "dry skin".

Although for ads from the same advertiser landing pages are different and even ads may differ, a similar template is mostly used for both the ad and the landing page. A model and corresponding features able to fit this regularity will be effective in learning the mobile friendliness and aesthetics of known (observed) advertisers and ads, but will have limited generalisability beyond a particular sample. Therefore, we split the data based on advertisers. We perform 10-fold cross validation and for each fold split advertisers in a group covering approximately 90% of the data and another group covering approximately 10% of the data.

Baseline: W3C mobileOK. There are other tools that analyze web pages and return a report describing what features of the page should be improved to increase their mobile friendliness. One of them is the W3C mobileOK library, an open source library that returns a mobile friendliness score between 0 and 100, in addition to a mobile friendliness analysis report, which has been used in previous work to analyse the mobile friendliness of sponsored search

Name	Description
error classes	POP-UPS, NON-TEXT-ALTERNATIVES, PAGE-TITLE, TABLES-NESTED, NO-FRAMES, PROVIDE-DEFAULTS, IMAGES-SPECIFY-SIZE, MEASURES, EXTERNAL-RESOURCES, STYLE-SHEETS-SUPPORT, CHARACTER-ENCODING-USE, PAGE-SIZE-LIMIT, CONTENT-FORMAT-SUPPORT, TABLES-LAYOUT, STYLE-SHEETS-USE, OBJECTS-OR-SCRIPT, VALID-MARKUP, MINIMIZE, IMAGE-MAPS, IMAGES-RESIZING.
critical	one hot enc of critical violations of error classes.
severe	one hot enc of severe violations of error classes.
medium	one hot enc of medium violations of error classes.
low	one hot enc of low violations of error classes.
warning	one hot enc of low violations of error classes.
numImageReq	number of loaded images.
numStyle	number of loaded stylesheets.
numDoc	number of time the page was loaded.
stylesheets	combined size of the stylesheets.
images	combined size of the images.
document	total size of the document.
score	mobileOK score.

Table 3: Features derived from the mobileOK library.

feature family	log reg	GBRT	GBDT	RF
readability	.531	.560	.597	.589
navigation	.599	.556	.631	.642
inputform	.582	.575	.681	.674
layout	.627	.585	.687	.706
texture	.641	.570	.704	.707
w3c	.687	.618	.722	.726
color	.663	.576	.712	.727
quality	.644	.618	.716	.727
windowsize	.687	.653	.724	.731
mobileoptimized	.697	.653	.745	.752
all	.761	.680	.777	.788

Table 4: Comparison of different models on the mobile friendliness classification task in terms of weighted AUC.

pages [21]. We aim to determine how the mobile friendliness recommendations of the mobileOK library (designed for general web pages) relate to the collected mobile friendliness judgements for sponsored search ad landing pages described in Section 4. We therefore encode the faults detected by the library and use them together with the mobileOK score as features to predict mobile friendliness. Table 3 has an overview of the features. These consist of 20 error classes that can have one of five levels, i.e., critical, severe, medium, low, and warning. We derive a total of 107 features from the W3C report, see [26] for more details on the W3C library.

5.2 Results

We first present the results of comparing different models on the mobile page friendliness and aesthetics prediction tasks. After selecting the best performing model we present detailed experiments with different feature categories for mobile page friendliness prediction and aesthetics prediction, respectively.

Model comparison. Table 4 shows the performance of our four models in terms of weighted AUC (AUC_w), defined as the average of the AUC scores for each class weighted by the support (number of true instances). This metric takes class imbalance into account. The parameters of each model are optimized via a grid search over each model parameter space. The best performing model is Random Forest (RF) with 200 full depth trees pruned to leaf nodes with 3 or more instances. It outperforms all other models except GBDT on two of the feature families. From now, we use RF.

Mobile friendliness. We present in Table 5 how different feature categories perform on the task of mobile friendliness prediction. We train the RF model using different categories of features. Performance is again measured using weighted AUC (AUC_w). We further break down the performance of the classifiers on each individual class, i.e., the AUC for each class versus the other classes. We observe that the model using all features outperforms any of the subsets of features representing different feature categories in terms of AUC_w. However, in terms of predicting the different classes of mobile friendliness versus the other classes, we observe some variation in the performance of different feature categories.

feature family	AUC_w	AUC_b	AUC_f	AUC_g	AUC_p
(r) readability	.589	.518	.662	.553	.586
(n) navigation	.642	.695	.677	.566	.626
(i) inputform	.674	.682	.697	.618	.661
(t) texture	.706	.747	.734	.662	.632
(l) layout	.707	.825	.722	.668	.620
(3) w3c	.726	.755	.752	.694	.725$^{\triangle}$
(c) color	.727	.795	.766	.676	.623
(q) quality	.727	*.855*	.719	.696	.627
(w) windowsize	.731	.850	.741	.707	.629
(m) mobileopt	.752	.797	*.769*	*.739*	.687
all	.788$^{\triangle}$	.890$^{\triangle}$	.800$^{\triangle}$	.762$^{\triangle}$	*.693*

Table 5: Weighted AUC as well as AUC scores per class as produced by the RF model. $AUC_{b,f,g,e}$ indicates the AUC for the specific class (bad, fair, good, perfect respectively) versus the others. $^{\triangle}$($^{\blacktriangle}$) indicates a significant improvement over the second highest score in the column at the $\alpha = .05$ ($\alpha = .01$) level. Italics indicate the second highest score in a column.

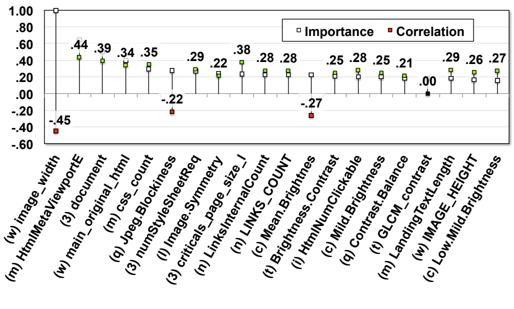
HTML based features		image based features	
feature category	AUC	feature category	AUC
inputform	.541	layout	.633$^{\triangle}$
navigation	.551	all	.661$^{\triangle}$
mobileopt	.566	quality	.680$^{\triangle}$
readability	*.570*	texture	.688$^{\triangle}$
		color	.695$^{\triangle}$

Table 6: AUC scores as produced by the RF model. $^{\triangle}$indicates a significant improvement over the highest scoring HTML based feature (in italics) at the $\alpha = .05$ ($\alpha = .01$) level.

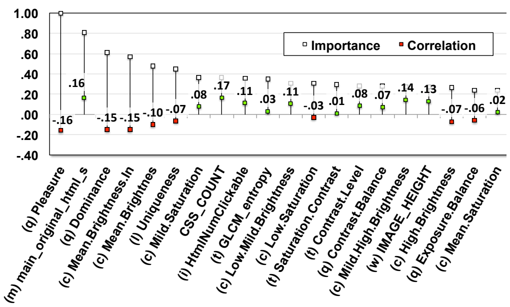
In differentiating between pages that are bad-or-not (AUC_b) we find that window size features are not as effective as all features combined but do achieve competitive performance We further observe that image quality features also perform well. With respect to fair-or-not as well as good-or-not, we observe in columns AUC_f and AUC_g, respectively, that *mobileopt* features achieve performance close to that of using all features combined. Features specific to mobile optimization, e.g., a click to call button, appear to be good indicators of whether a page is fair or good.

Regarding mobile pages judged as perfect-or-not we find that the W3C mobileOK features are more effective than using a combination of all other features. We take a closer look at the importance of specific features in the next section.

Aesthetic appeal. A comparison of the effectiveness of HTML and image based feature categories on the task of aesthetics prediction is presented in Table 6. We observe that HTML based features are bad predictors for the aesthetics of a mobile page. Image based features, however, perform significantly better and especially color based features, i.e., reaching a performance of .695 AUC, which is in line with state-of-the art results on image beauty classification in computational aesthetics [22]. The combination of image and HTML based features does not appear to contribute to a more discriminative model and the performance of using a combination of all features does not significantly differ from that of any of the image feature categories individually.



(a) Top 20 features for mobile friendliness



(b) Top 20 features for aesthetics

Figure 1: Feature correlation coefficient and importance for the top 20 most discriminative features (feature family abbreviation) for mobile page friendliness (left) and aesthetics (right). See Table 5 for feature family abbreviations.

5.3 Analysis

We aim to answer the following questions: "What makes ad landing pages mobile friendly?", "What makes ad landing pages aesthetically pleasing?", "How do aesthetically pleasing pages differ from mobile friendly pages?". We look at the importance of individual features for mobile friendliness and aesthetics prediction. We combine the best performing feature sets (all and W3C) in a single model and compute, for each feature, the following metrics:

1. **Feature Importance** Given the effectiveness of the RF model built in Section 5.1, we inspect which features such a classifier considers more discriminative to distinguish between different classes of mobile friendliness. We use out of bag permutations to identify the importance of the features.
2. **Feature Correlation** To investigate the *direction* of the association between important features and dependent variables, we perform a correlation analysis and compute Spearman's *rho* between each feature and the ad labels. For aesthetics we also use Spearman's correlation between each feature and the aesthetics judgements.

We report results for the top-20 most important features in Figure 1, for both mobile friendliness and aesthetics prediction. Finally, to compare aesthetics with mobile friendliness, we compare correlations of the most important features for mobile friendliness and aesthetics prediction, and report results in Figure 3.

Mobile Friendliness. The most discriminative feature is *image width*, i.e., the width of the rendered HTML image. Two other window size features, *main original HTLM size* and *image height* are also important. These features are computationally inexpensive and effective. Other important features include optimized features, such as *MetaViewportExists*, *css count*, and *landingPageTextLength*) and the W3C family features (*document size*, *numStyleSheetsRequested*, and *criticals page size limit*), which focus on the amount of stylesheets used in a page and the page size. These features capture the size of the page and amount of stylesheets affecting the performance of mobile landing pages.

The most discriminative image based feature for mobile friendliness prediction is the *JPEG blockiness*, i.e., the amount of compression noise in the image. Since there is no compression in the image of the rendered HTLM, JPEG blockiness becomes a measure of overall page smoothness. This feature is negatively correlated with mobile friendliness, thus suggesting that images of pages that look cleaner and less pixelized tend to be more mobile friendly.

As noted in previous work on web aesthetics [2], shape and color *balance* is crucial for usable interfaces. We indeed observe that both *Symmetry* and *Contrast Balance* are positively correlated with mobile friendliness. Figure 2 shows examples of ad landing pages with high contrast balance and symmetry.

We find that low quality mobile pages more often have only very bright color combinations, whereas for high quality mobile friendly pages the average *Mean Brightness* tends to belong to the lower range (*Low/Mild*). We also see a positive correlation with *Brightness Contrast* indicating that diversity in brightness values is more often observed in high quality mobile friendly pages, i.e., the presence of a small number of bright colors with an otherwise moderate use of brightness tends to be associated with mobile friendly pages.

To characterise the mobile friendliness of ad landing pages we run the model on a sample of ad landing pages captured during one week. Using only the HTML based features, we find that 6% of the landing pages in our sample are bad in terms of mobile friendliness. Most ad landing pages are classified as 50% fair, suggesting that some mobile optimization is present but that pages are otherwise lacking in ad mobile friendliness. The remaining 35% is good, and 9% of the mobile pages are rated as perfect. The latter figure is particularly surprising as it suggests that the experience provided by advertisers on their landing pages is not helping users engaging with the advertised product and converting (e.g. making a call).

We gather conversion and dwell time data during one month for the ads in our above sample and perform a χ^2 test to assess whether the distribution of impressions with and without conversions differ over the four classes. We find a significant difference between the distribution of conversions over the four classes (χ^2=453.4, pval$<$0.0001). We inspect the adjusted residuals and find that the perfect mobile friendly class has a residual of -4.5 indicating that less conversions are observed in this class than expected. In contrast the bad mobile friendliness class has a residual of 10.1 indicating that more conversions are observed than expected. Although surprising, conversion rates have been found to differ considerably depending on the type of landing page [3]. Such variations may explain our observations; consider for example a conversion defined as a page visit versus a conversion defined as an email sign-up or buying a product. However, we do find an increase in clicks with high dwell time (long clicks), an indication of positive post-click user experience [19], (χ^2=1908.2, p-val$<$0.0001) with standardized residuals of 13.2 for the perfect class of ad landing pages and -8.9 for the bad class.

Aesthetic Appeal. The analysis in Figure 1 reveals that, unlike mobile friendliness, aesthetic appeal of landing pages can be mostly explained with visual features. This is expected; aesthetic assessment involves the evaluation of the page visuals only, without considering page usability or efficiency. More specifically, we notice that most of the discriminative features for this task relate to the notion of image *Brightness*. The first most important such visual feature are *Pleasure* and *Dominance*. The negative correlation between these features and mobile page aesthetics is due to the fact that these descriptors, originally designed for natural photographs,

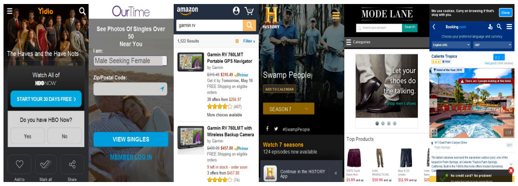


Figure 2: Examples of mobile friendly pages judged as perfect that were also judged to be aesthetically pleasing

are directly proportional to the amount of brightness in the picture. As said, pages that are too bright might annoy the user (see negative correlation between *High Brigtness* and aesthetics), while pages with *Mild Brightness* tend to be more aesthetically pleasing. Also, a wide range of brightness shades (high *Contrast Balance*), is preferred over a less exciting *Exposure Balance*. Generally, color saturation is also positively related to aesthetic appeal.

Figure 2 shows six examples of mobile friendly pages judged as perfect *and* aesthetically pleasing. The use of brightness in these images is limited to certain buttons or areas to call the users attention but otherwise presents a smooth and balanced design. Overall, the strong presence of the above mentioned visual features as discriminatory factors suggests that designers should carefully look at the perceived brightness of the landing page to create aesthetically pleasing ads.

In terms of composition, unlike creative media or beautiful pictures [29], the uniqueness of the object arrangement is inversely correlated with page beauty, whereas the presence of objects in certain focus areas is a positive indicator of page aesthetic appeal. This suggests that aesthetically pleasing mobile friendly pages follow a particular pattern and that deviating from this pattern is not associated with mobile friendly or aesthetic pages.

To characterise the aesthetics of ad landing pages, we run the model on a sample of ad landing pages captured during one day. Further conversion data is extracted over one month for the ads in the sample. We find a significant difference between the observed and expected conversions for the aesthetics classes (χ^2=247.8 p-val$<$0.0001) with the aesthetically appealing class having an adjusted residual of 26.9 compared to -1.0 for the non appealing class. This suggests that aesthetically pleasing pages attract more conversions.

Mobile Friendliness vs. Aesthetics. In our dataset, the correlation between mobile friendly and aesthetics scores is 0.27. To understand the reasons for such low correlation, we compare in Figure 3 correlations of the most important features for mobile friendliness and aesthetics classification. Many interesting patterns arise.

From a visual feature perspective, we find that, unlike mobile friendliness, the smoothness of landing pages (*Jpeg.Blockiness*) does not affect their aesthetic appeal, thus suggesting that both smooth and rougher pages can be seen as pleasing. Similarly, whereas page *Symmetry* is very important for usability and mobile friendliness, it does not represent a discriminative factor for aesthetic appeal. Although we saw that high *Brightness* can be annoying and cause low friendliness and aesthetic appeal, the importance of brightness for aesthetics is less prominent compared to mobile friendliness. Similarly, low uniqueness is more important for mobile friendliness than for aesthetics.

Regarding the HTML features we find that features related to whether the mobile landing page is optimized (properly rendered when on a mobile device) or structured in a way that is helpful for users to consume are important to identify mobile friendliness, but much less so for aesthetic appeal. Interestingly, the width of

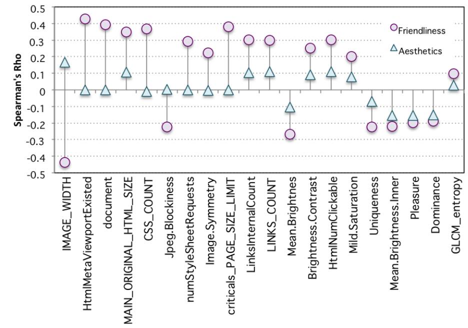


Figure 3: Correlation-based comparison between the most important features for mobile friendliness and aesthetics prediction.

the image is negatively correlated (high) to mobile friendliness, but positively correlated (low) to aesthetics.

Overall these results show that mobile friendliness and aesthetic are two different, albeit related, dimensions, of the quality of *ad* landing pages. Accounting for both is likely to not only provide easier navigation, but also a more inspiring and convincing experience, which is particularly important in mobile advertising.

6. CONCLUSIONS

Our aim was to study two dimensions of the post-click experience of landing pages in the context of sponsored search, so far largely unexplored, namely mobile friendliness and aesthetic appeal. Although the study of mobile friendliness is not new, and there are tools that can be used by interested parties to check how mobile friendly their pages are, this work is the first that provides a systematic study of what makes an "ad" landing page mobile friendly. In particular, our dataset is one of the first and largest high quality manually crafted datasets characterizing mobile friendliness in the context of advertising. It accounts for the post-click experience in terms of the user being able to "consume" the ad.

We found that less than half of the ad landing pages are good or perfect in terms of mobile friendliness and that mobile friendly ads are positively associated with long clicks. However, we do not find that such ads lead to more conversions. Regarding aesthetic appeal we found that few ads are visually attractive, which is surprising in the context of advertising, where the aim is to project a good image of the brand and/or product(s) being advertised. The literature in web design and user engagement even emphasizes the aesthetics of landing pages. Our data seems to confirm this as aesthetically appealing pages are positively associated with conversions.

With respect to predicting mobile friendliness we found that a model using a mix of visual and structural features is most effective, while visual features are most important for predicting aesthetic appeal. Also, the W3C recommended mobileOK checker has good but nonetheless limited ability to predict most categories of mobile friendliness. The potential of visual features and aesthetic appeal suggests that their use should increase in tools developed to assess page mobile friendliness and the quality of the user experience.

Finally, we recently launched an experiment, where we use the model described in this paper with the set of HTML and visual features, to filter out the bad ads in terms of mobile friendliness. Initial results show a significant decrease (-6.04%) of short clicks, which are clicks with short dwell time reflecting a negative post-click user experience [19]. We also saw an increase of 1.15% in ad click-through rate, suggesting that bad landing pages are bad ads in general, echoing results in [1, 19].

References

[1] N. Barbieri, F. Silvestri, and M. Lalmas. Improving post-click user engagement on native ads via survival analysis. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016*, pages 761–770, 2016.

[2] M. Bauerly and Y. Liu. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies*, 64(8):670–682, 2006.

[3] H. Becker, A. Broder, E. Gabrilovich, V. Josifovski, and B. Pang. What happens after an ad click?: Quantifying the impact of landing pages in web advertising. In *Conference on Information and Knowledge Management*, 2009.

[4] H. Becker, A. Broder, E. Gabrilovich, V. Josifovski, and B. Pang. Context transfer in search advertising. In *SIGIR*, 2009.

[5] H. Cheng, R. v. Zwol, J. Azimi, E. Manavoglu, R. Zhang, Y. Zhou, and V. Navalpakkam. Multimedia features for click prediction of new ads in display advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 777–785, 2012.

[6] Y. Choi, M. Fontoura, E. Gabrilovich, V. Josifovski, M. Mediano, and B. Pang. Using landing pages for sponsored search ad selection. In *World Wide Web*, 2010.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR, pages 886–893, 2005.

[8] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*, volume 3953 of *Lecture Notes in Computer Science*, pages 288–301, 2006.

[9] A. De Angeli, A. Sutcliffe, and J. Hartmann. Interaction, usability and aesthetics: What influences users' preferences? In *DIS '06*, pages 271–280, 2006.

[10] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621, 1973.

[11] H. Hohnhold, D. O'Brien, and D. Tang. Focusing on the long-term: It's good for users and business. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1849–1858, 2015.

[12] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *CVPR*, pages 1–8. IEEE, 2007.

[13] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR, pages 145–152, 2011.

[14] Y.-G. Jiang, Y. Wang, R. Feng, X. Xue, Y. Zheng, and H. Yang. Understanding and predicting interestingness of videos. In *AAAI*, 2013.

[15] A. Kae, K. Kan, V. K. Narayanan, and D. Yankov. Categorization of display ads using image and landing page features. In *LDMTA*, 2011.

[16] Y. Ke, X. Tang, and F. Jing. The design of high-level features for photo quality assessment. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 419–426, 2006.

[17] A. Khosla, A. Das Sarma, and R. Hamid. What makes an image popular? In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 867–876, 2014.

[18] M. Lalmas, H. O'Brien, and E. Yom-Tov. Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 6(4):1–132, 2014.

[19] M. Lalmas, J. Lehmann, G. Shaked, F. Silvestri, and G. Tolomei. Promoting positive post-click experience for in-stream yahoo Gemini users. In *SIGKDD'15*, pages 1929–1938, 2015.

[20] T. Lavie and N. Tractinsky. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.*, 60(3):269–298, 2004.

[21] H. Liu, W.-C. Kim, and D. Lee. Characterizing landing pages in sponsored search. In *LA-WEB*, 2012.

[22] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang. Rapid: Rating pictorial aesthetics using deep learning. In *Proceedings of the ACM International Conference on Multimedia*, pages 457–466, 2014.

[23] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92, 2010.

[24] K. Mo, B. Liu, L. Xiao, Y. Li, and J. Jiang. Image feature learning for cold start problem in display advertising. In *IJCAI'15 Machine Learning Track*, 2015.

[25] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato. Aesthetic quality classification of photographs based on color harmony. In *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR, pages 33–40, 2011.

[26] S. Owen and J. Rabin. W3c mobileok basic tests 1.0. *W3C Candidate Recommendation*, 30, 2007.

[27] K. Oyibo, Y. S. Ali, and J. Vassileva. An empirical analysis of the perception of mobile website interfaces and the influence of culture. In *Proceedings of the International Workshop on Personalization in Persuasive Technology (PPTâĂŹ16), Salzburg, Italy*, 2016.

[28] F. Qian, S. Sen, and O. Spatscheck. Characterizing resource usage for mobile web browsing. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pages 218–231, 2014.

[29] M. Redi, N. O?Hare, R. Schifanella, M. Trevisiol, and A. Jaimes. 6 seconds of sound and vision: Creativity in micro-videos. *Computer*.

[30] M. Redi, D. Quercia, L. T. Graham, and S. D. Gosling. Like partying? your face says it all. predicting the ambiance of places with profile pictures. *arXiv preprint arXiv:1505.07522*, 2015.

[31] K. Reinecke, T. Yeh, L. Miratrix, R. Mardiko, Y. Zhao, J. Liu, and K. Z. Gajos. Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2049–2058, 2013.

[32] R. Schifanella, M. Redi, and L. M. Aiello. An image is worth more than a thousand favorites: Surfacing the hidden beauty of flickr pictures. In *ICWSM'15: Proceedings of the 9th AAAI International Conference on Weblogs and Social Media*. AAAI, 2015.

[33] D. Sculley, R. G. Malkin, S. Basu, and R. J. Bayardo. Predicting bounce rates in sponsored search advertisements. In *knowledge discovery and data mining*, 2009.

[34] E. Sodomka, S. Lahaie, and D. Hillard. A predictive model for advertiser value-per-click in sponsored search. In *International Conference on World Wide Web*, 2013.

[35] O. Starov, S. Vilkomir, A. Gorbenko, and V. Kharchenko. Testing-as-a-service for mobile applications: state-of-the-art survey. In *Dependability Problems of Complex Information Systems*, pages 55–71. Springer, 2015.

[36] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–477. IEEE, 2002.

[37] X. S. Zheng, I. Chakraborty, J. J.-W. Lin, and R. Rauschenberger. Correlating low-level image statistics with users-rapid aesthetic and affective judgments of web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–10, 2009.

[38] C. Zhong, D. Karamshuk, and N. Sastry. Predicting pinterest: Automating a distributed human computation. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1417–1426, 2015.

[39] K. Zhou, M. Redi, A. Haines, and M. Lalmas. Predicting pre-click quality for native advertisements. In *Proceedings of the 25th International Conference on World Wide Web*, pages 299–310. 2016.