

User Motion Prediction in Mobile Multimedia Application

Genyuan Zhang

The Zhejiang University of Media and Communications, Hangzhou, P.R. China

zgenyuan@163.com

Keywords: Multimedia; User Motion; Prediction;

Abstract. The paper describes an approach to user motion in mobile multimedia based primarily on features. The increasing number of mobile display is based on HTML5 web standard. The openness of this standard makes it easier to extract some additional features, such as animation, video, sound and other components. Using these features in conjunction with global image features, user features and publisher features to improve user motion.

Introduction

Mobile multimedia is a relatively new, fast growing market. According to Gartner, mobile multimedia revenue will grow from \$9.6 billion in 2012 to \$11.4 billion in 2013. By 2016, mobile multimedia revenue is expected to reach \$24.5 billion. While mobile search drives a large portion of that revenue, Gartner believes mobile display ad spending will eventually take over [1]. Display multimedia on mobile devices generates revenue by showing graphical ads either on web pages or inside applications. Similarly to how display work on traditional computers, users of mobile devices, such as smart phones or tablets, are shown a relatively small banner unit. If they click on the banner, users are taken either to a landing page or an expanded ad unit, which is where a more profound engagement with the advertiser's brand can occur. Display ad clicks drive revenue for advertisers, usually through increased user engagement with their brand. They are also revenue drivers for publishers, multimedia networks and other members of the mobile multimedia ecosystem, whose revenue is also influenced by user engagement. This influence is sometimes direct, which is the case with cost-per-click (CPC) or cost-per-action (CPC) pricing models. This is the main reason why different user motion and ad optimization methods have been used to try to match banner ads with users who are most likely to click them. Display multimedia on mobile devices is technically different to that on desktop computers. Users interact with mobile devices differently and in different contexts. Location plays a much more important role, while accelerometer and touch interface enable new ways of interaction. Rich media ads, with features such as animation, sound or video, have commonly been built using Flash. Since Flash is either unavailable or underperforming on mobile devices, HTML5 has been adopted as the defacto standard for rich media mobile multimedia[2].

Related Work

In large part, our work builds on multimedia feature extraction done by Cheng et. al. at Yahoo! Labs [3]. We used several global image features from their work. While their data includes banner ads as static images or Flash files, our data includes banner ads in HTML5 format, which allowed us to extract some additional ad features. Most work in user motion in online multimedia focused either on Contextual Multimedia(Context Match, CM) [6] or Sponsored Search (SS) [7]. In both of these areas, the goal is to find a match between an ad, which is observed as a textual document, and a query, which can either be a web page or a search query. Given the reliance on textual features, most approaches used in CM or SS cannot be applied directly in display advertising. While image features have been researched in the field of content based image retrieval (CIBR), we haven't been able to find much work, that correlates those features to click-through rates in online advertising. The work of Azimi et. al. [4] and Cheng et. al. [3] are among the first to suggest using multimedia features for estimating clicks in display advertising. In [4], a relatively small dataset was used and it

included only static images. The main contribution of our work is in expanding on [3] by introducing a novel ad representation incorporating HTML5-based features.

User motion

The aim of our work was to estimate how likely it is that a given banner ad will be clicked in a given context. The dataset we used was obtained from mobile multimedia that were served by Celtra in December 2012 and January 2013. The dataset contains sessions for 1076 different expandable banner ads. Expandable banner ad format is an HTML5-based ad format that consists of two ad units: a banner unit and an expanded unit. The banner unit, which is usually relatively small, is shown on a publisher's web page or inside a publisher's application. Once the banner unit is tapped (equivalent to a click on traditional computers), the larger expanded unit is shown. Each session in the dataset represents one expandable banner ad impression and all consecutive interactions between the user. Given our goal to predict banner clicks, we were only interested in data related to the structure and interactions on the banner unit. The features were extracted from three sources: tracking logs, ad metadata and banner screenshots. Tracking logs contain events that happen while campaigns are running and ads are being served. Sessions, which contain user and publisher features, as well as timestamps and information whether the banner was clicked, were extracted from the logs. HTML5 features, such as size, the presence of animation, the number of different components and the number of actions, were extracted from ad metadata, which was retrieved from API. Image features, such as brightness, saturation, colorfulness, contrast, naturalness and hue, were extracted from banner screenshots, obtained by rendering each of the banner ads in a headless browser. All numeric features were discretized based on histograms derived from the dataset. Features used to represent an ad are obtained from three sources: ad metadata, banner screenshots (image) and tracking logs (session). Metadata features are obtained as creative and HTML5 features as follows:

- Area is the surface area of the banner in pixels, so that the area equal or below 50,000 pixels is considered small, while the banner ads with area above that threshold are considered large.
- Aspect ratio is the ratio of the width of a banner ad to its height. 1 - 4 is considered narrow, 5 and 6: medium and 7 or higher: wide.
- Device type is either phone or tablet. In our particular case, the device type to which an ad will be served is predetermined, which is why we consider it a creative feature, rather than a session feature.
- Animation distinguishes between banners that are static and banners that are animated using CSS3-based animation properties [8].
- Number of components represents the number of top-level HTML elements that are visible. These are usually images but can also be video, buttons, form components or more complex components, such as maps, galleries, etc. Discretization was such that 1 or 2 components were considered low, 3 to 8: medium and 8 or more: high.
- Number of actions represents the number of specific JavaScript function calls, which trigger some behavior, such as animation. Actions are triggered by events, the most common of which is the "appear" event, which fires when a banner is loaded and shown to the user. The most common actions on the load event are animation actions, so the number of actions usually correlates to the complexity of animation on the banner. Upon discretization, we considered less than 30 actions as low, 30 to 70 actions as medium and 70 or more actions as high.

Banner screenshots are used as a source of image features as follows:

- Brightness of an image was derived from the HSL color space [9], where "L" measures the lightness of color. It is 0 for white and 1 for black, whereas all other colors are somewhere in between. The mean brightness of pixels in the image and the standard deviation of all pixels in the image were extracted as brightness features.
- Saturation measures the vividness of an image, the value of which was derived from the HSL color space. Similarly to brightness, we used mean and standard deviation as features.
- Colorfulness is a measure of distance against gray, which is calculated in RGB [10].

- Contrast measures the relative variation of luminance across the image in HSL color space.
- Naturalness is the degree of correspondence between images and human perception of reality, based on the approach proposed by Huang et al. [10].

- Hue value from the HSL color space was used to determine the predominance of a certain color in an image. Mean hue value and standard deviation were used as features.

The user tracking logs are used as a source of session features as follows:

- Platform represents the operating system installed on the user's device. It is either iOS or Android and a version number.

- In-app is the feature that distinguishes between banner impressions on web pages and inside applications. Unlike in display multimedia on traditional computers, where banner ads are shown on web pages only, in mobile display advertising, a large number of banner impressions come from applications installed on mobile devices.

- SDK is a software component that displays an HTML5 banner ad inside an application. Different SDKs are implemented using different technologies and support different ad features. Different SDKs also imply different publishers.

- Referrer is a string retrieved from the HTTP header, set by the publisher web page or application that displayed the banner ad. This string is commonly used to identify the publisher.

For the purpose of evaluation, we have estimated the model parameters from the training data (obtained in December 2012) and evaluated the performance of the model on the test data (obtained in January 2013). We estimate clickthrough rates for banner ads with more than 50,000 impressions in January and then compare that to the actually observed value. The best performing feature set in most percentiles is a combination of all three feature types: creative/HTML5, image and session. The features in this set are: animation, area, aspect ratio, device type, number of actions, number of components, hue mean, hue standard deviation, saturation mean, saturation standard deviation, brightness mean, brightness standard deviation, contrast, colorfulness, naturalness, SDK.

Conclusion

This paper proposes a novel representation for mobile multimedia with the aim of motion prediction. Adding metadata (creative and HTML5) features to image and session features can clearly improve predictions. In our future work, we plan to expand the feature sets in order to incorporate the context. The session data we currently have could be extended with data about the user, such as location, gender, age and other demographic data.

References

1. S. Baghdassarian and A. Frank. Forecast: Mobile Advertising, Worldwide, 2009-2016. Gartner, 2012.
2. R. Paley. HTML5 - bridging the mobile platform gap: mobile technologies in scholarly communication. In *Serials: The Journal for the Serials Community*, 2011.
3. H. Cheng, R. van Zwol, J. Azimi, E. Manavoglu, R. Zhang, Y. Zhou, and V. Navalpakkam. Multimedia features for user motion of new ads in display advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 777-785. ACM, 2012.
4. J. Azimi, R. Zhang, Y. Zhou, V. Navalpakkam, J. Mao, and X. Fern. The impact of visual appearance on user response in online display advertising. *CoRR*, 2012.
5. M. Zaffalon and M. Hutter. Robust feature selection by mutual information distributions. In *18th International Conference on Uncertainty in Artificial Intelligence*, pages 577 - 584, 2002.
6. A. Broder, M. Fontoura, V. Josifovski, and L. Riedel. A semantic approach to contextual advertising. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 559-566. ACM, 2007.

7. V. Abhishek and K. Hosanagar. Keyword generation for search engine multimedia using semantic similarity between terms. In Proceedings of the ninth international conference on Electronic commerce, pages 89 - 94, New York, NY, USA, 2007. ACM
8. G. Williams. CSS3 for Mobile. Learn HTML5 and JavaScript for Android (2012): 119-156.
9. G. H. Joblove and D. Greenberg. Color spaces for computer graphics. Computer Graphics (SIGGRAPH '78 Proceedings), 12(3):20 - 25, Aug. 1978.
10. D. Hasler and S. E. Suesstrunk. Measuring colorfulness in natural images. Human Vision and Electronic Imaging VIII, Proceedings of the SPIE, 5007:87 - 95, 2003.