2019 IEEE International Conference on Consumer Electronics
- Taiwan (ICCE-TW)

Cross-Platform Measurement on Ad Exchanges

Chenyue Zhang, Chunxi Li, Yongxiang Zhao, Nanxi Huang
School of Electronic and Information Engineering,
Beijing Jiaotong University, Beijing, China 100044
{18120172, chxli1, yxzhao, 15120204}@bjtu.edu.cn

Baoxian Zhang
University of Chinese Academy of Sciences,
Beijing, China 100049
bxzhang@ucas.ac.cn

Abstract—**As a key component of online advertising ecosystem, Ad exchange platform (ADX) has been studied extensively. However, existing work is mainly for studying individual ADX platform, and they are not suitable for measuring and comparing the performance of different ADXs. A challenge for comparing the performance of multiple ADXs is as follows: How to enable different ADXs to acquire the same piece of information from a same user, such that we can accurately compare their performance according to the ads they delivered to the same user. For this purpose, in this paper, we propose a parallel cross-platform measurement method, which creates a user (a virtual role) to access the webpages jointly monitored by different ADXs to enable these ADXs to acquire the user-profile from the same user. We collect ads that the ADXs deliver to same user, and compare their ad targeting performance based on these ads. The results validate the efficiency of our proposed method.**

I. Introduction

In an online ad ecosystem, Ad exchange platform (ADX) is a key component bridging customers and advertisers. It tracks user data, taps user interest, and pushes advertisements to users according to their interests. Since ADX usually makes final decision on which ads could be displayed and to which users via ad targeting and thus affects the revenue of both the advertisers and the ADX platform, it is of special importance for multilateral participants in online ad ecosystems.

There have been many works focusing on evaluating ad targeting ability of ADX. Refs. [1, 2] measured Google ad targeting ability from the perspective of web users, and confirmed that Google serves search engine ads and general webpage ads according to different users' search history [1] and different users' personas [2], respectively. Ref. [3] harvested ads from webpages crawled with profile-based crawler and found that ad targeting is widely adopted by different ADX platforms.

Moreover, due to a large number of ADXs with competitive relationship in the ad ecosystem, distinguishing the performance between ADXs is an interesting problem worth to be studied. However, most existing work focused on single ADX. They are not suitable for comparative study of multiple ADXs.

Different from single-ADX-based research, cross-ADX-platform research faces a key challenge: We need to let the multiple ADXs acquire the same user information, relying on which these ADXs can target ads toward the user, such that the ad targeting performance difference between ADXs can be evaluated accurately due to the performance difference mainly come from the ADX side.

This work was supported in part by National Natural Science Foundation of China under Grants 61872031, 61572071, 61872331 and U1534201.

In this paper, we propose a parallel cross-platform measurement method for comparing ad targeting ability of different ADXs. The key idea is to find the pages monitored jointly by multiple ADXs, and generate training-page sets from these pages so as to shape the same user (virtual role) simultaneously on these ADXs by accessing the training-pages. During the measurement process, we access some non-training pages to collect displayed ads, which will be used for performance evaluation of the ADXs.

II. Parallel Cross-platform Measurement Method

In this section, we propose a cross-platform measurement method to compare multiple ADXs' performance.

A. Find intersection pages

First, we need to find pages (called intersection pages) monitored jointly by these ADXs and generate training-page sets from these intersection pages, for shaping the same virtual role simultaneously on the ADXs, in order to accurately capture the difference between different ADX platforms. For this purpose, we use OpenWPM [4], an automated web privacy measurement platform developed by Princeton, to collect data of http requests, http responses, and third-parts' embedded links, by accessing top n webpage URLs ranked by Alexa (alexa.com). Based on the collected data, we find the webpages monitor by an ADX, and further extract the webpages jointly monitored by multiple ADXs to form intersection pages.

B. Generate training page set

In this step, we need to generate training-page sets from the intersection pages, for shaping the same virtual role with different traits simultaneously on the ADXs. First, for each web page, it performs word segmentation after removing non-Chinese characters [5], and uses TF-IDF to get the top 20 feature words to represent the page. Then, for all web pages, it uses the k-means to classify the intersection pages into different clusters as training-page sets, and uses the Silhouette Coefficient to find an appropriate k value [6]. Finally, comparing a cluster's main keywords with ad keywords published by Google AdWords, it links each cluster to an ad category closest to it and uses the ad category name as cluster tag, to facilitate subsequent ADXs' performance evaluation.

C. Train virtual role and Extracting ads

Similar to [2], virtual roles rely on different user-agent settings, and for a virtual role with certain trait, the role training and the ads extracting are simultaneously performed. Specifically, it uses a user-agent to randomly access pages in a training-page set with certain tag to train a virtual role with a trait, and at the same time to randomly access some non-training pages (called control pages) to trigger ad ADX targeting toward the virtual role. We use typical portal pages as control pages, which have no obvious interest bias and have enough ad placements for different ADXs. Therefore, accessing them should cause little impact on the virtual role's trait and can obtain enough targeted ads. Moreover, since direct accessing to ad pages with virtual role affects its traits, we use a particular routine instead the virtual role itself to access the ads URLs and open landing pages, to acquire the ads contents.

D. Mine ad keywords

We mine ads keywords from the acquired ads contents, in order to quantitatively compare ad-targeting difference between ADXs. We adopt three methods for ads keywords mining. Assume a given landing page. If it is a text HTML file, we try to extract keywords from the title or the product description; If it is just an ad image, we use Google Image Search Engine to help parse image and get keywords; Finally, if these two methods do not work, we send landing page's URL to Google AdWords keyword planner to get keywords. Applying these methods, we get keywords corresponding to each landing page, with which we can perform performance evaluation in the next section.

III. Performance Evaluation

A. Datasets and key performance indicators

We found 10,355 Chinese pages monitored by more than 9 ADXs, from the Alexa's top one million websites. Among them, Google or Baidu ADX monitored 5,214 and 4,224 pages, far outnumbering any other ADXs. Thus, we measured Google and Baidu ADXs in Jan. 2018 for evaluating our method. Specifically, we used the webpages monitored by both ADXs to generate different training-page sets, each for training a specific trait for a virtual role on the two ADXs, and let each such virtual role randomly access some control pages to collect displayed ads for performance evaluation.

We use two key performance indicators (KPIs) as proposed in [2] to characterize an ADX's ad targeting: Targeted Training Keywords (TTK) and Behavioural Advertising in Landing Pages (BAiLP). For a given role trait i, TTK value T_i represents the training page keywords displayed on the landing page as a percentage of the total training page keywords; BAiLP value B_i reflects the percentage of targeted ads in all the ads received by a virtual role i. In contrast, T_i is more sensitive than B_i, since the displayed ads usually have a larger keyword set than the training pages.

B. Numerical results

1) Static behavior of ads targeting

We created three virtual roles, which have traits of military, finance, and design, respectively, and used them to measure Google and Baidu ADXs. We compute a TTK value for each role on either ADX, and show all the values in Fig. 1, where each corner corresponds a specific role trait. Fig. 1 indicates that, in general, both ADXs can target ads to roles based on different role traits, but have different targeting strengths. It shows that the ads toward finance and design roles are significantly targeted on both ADXs, as we have TTK values for (finance role, design role) equal to (1.0, 1.0) and (1.0, 0.75) for Baidu and Google ADXs, respectively. This is partially in line with the finding in [2], which reported a strong ad targeting for financial personas on Google ADX. In contrast, for the military role, targeting is not apparent: Baidu did not targeted ads, and Google targeted few ads toward this role, which was not reported in the previous literatures.

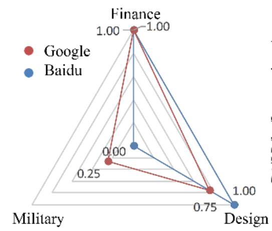

Fig. 1. TTK of both ADXs.

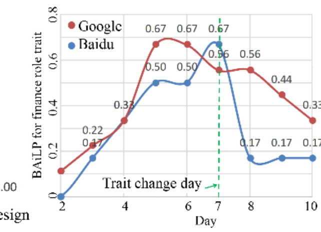

Fig. 2. BAiLP of both ADXs.

2) Dynamic behavior of ads targeting

We next measure the ad targeting ability of Google and Baidu ADXs on a virtual role with varying trait. The measurement lasted for 10 days, during which role trait was set to finance and education in the first 6 and last 4 days, respectively. We extracted keywords from the training and the landing pages and calculated BAiLP per day for each ADX.

Fig. 2 shows the results, where the x- and y-axis represent the sequence number of measurement day and the BAiLP value for finance trait, respectively; The vertical dash line indicates the role trait change day. From Fig. 2, it can be seen that, in contrast to Baidu, Google is more sensitive to finance trait and has higher average intensity of ad targeting toward this role. Google started to target ads from day 2, one day earlier than Baidu's start time; Google reached the maximum intensity (BAiLP, 0.67) of ad targeting on day 5, two days earlier than Baidu reaching the same intensity. After the trait change day (day 7), Baidu's targeted ads toward the original finance trait sharply reduces to a very low level (0.17), while Google's curve decays rather slowly. It implies that Google remembered the finance role trait more persistently than Baidu.

References

[1] S. Guha, B. Cheng, and P. Francis. "Challenges in Measuring Online Advertising Systems," in *Proc. of ACM IMC,* pp. 81-87, Melbourne, Australia, Nov. 2010.

[2] J. Mikians, J. Mikians, V. Erramilli, V. Erramilli, and N. Laoutaris, "I Always Feel Like Somebody's Watching Me: Measuring Online Behavioural Advertising," in *Proc. of ACM CoNEXT '15*, pp. 1-13, Heidelberg, Germany, Dec. 2015.

[3] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, "Adscape: Harvesting and Analyzing Online Display Ads," in *Proc. of WWW '14*, pp. 597-608, Seoul, Korea, Apr. 2014.

[4] S. Englehardt and A. Narayanan, "Online Tracking: A 1-million-site Measurement and Analysis," in *Proc. of ACM CCS '16*, pp. 1388-1401, Vienna, Austria, Oct. 2016.

[5] Jieba Chinese text segmentation, https://github.com/fxsjy/jieba

[6] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley-Interscience, Mar. 2005.