Rethink Targeting: Detect ‘Smart Cheating’ in Online Advertising through Causal Inference

Pengyuan Wang† Dawei Yin† Marsha Meytlis§ Jian Yang† Yi Chang†
†Yahoo Labs, Sunnyvale, CA, US §The Weather Channel, New York, NY, US
†{pengyuan, daweiy, jianyang,yichang}@yahoo-inc.com §masha707@gmail.com

ABSTRACT

In online advertising, one of the central questions of ad campaign assessment is whether the ad truly adds values to the advertisers. To measure the incremental effect of ads, the ratio of the success rates of the users who were and who were not exposed to ads is usually calculated to represent ad effectiveness. Many existing campaigns simply target the users with high predicted success (e.g. purchases, searches) rate, which often neglect the fact that even without ad exposure, the targeted group of users might still perform the success actions, and hence show higher ratio than the true ad effectiveness. We call such phenomena ‘smart cheating’. Failure to discount smart cheating when assessing ad campaigns may favor the targeting plan that cheats hard, but such targeting does not lead to the maximal incremental success actions and results in wasted budget. In this paper we define and quantify smart cheating with a smart cheating ratio (SCR) through causal inference. We apply our approach to multiple real ad campaigns, and find that smart cheating exists extensively and can be rather severe in current advertising industry.

Categories and Subject Descriptors: G.3 Probability and Statistics: Statistical Computing; J.1 Administrative Data Processing: Business, Marketing

Keywords: Advertising, Targeting, Causal Inference

1. INTRODUCTION

One of the major purposes of advertising is to draw extra success actions and hence additional revenue for the advertisers. To measure the incremental effect of ads, the ratio of the success rates of the users who were and who were not exposed to ads is usually calculated to represent ad effectiveness. Such a ratio is referred to as a **naive amplifier** of the ad [3] since it is supposed to represent the ad’s amplifying effect. However, the naive amplifier can be falsely inflated by targeting the users with high success intention even without ad exposures, and hence results in wasted budget. For example, consider an ad campaign for a cosmetic product that is already popular among teenage females. The campaign is designed to show impressions to teenage females who would regularly buy the product regardless of ad exposures. Such a campaign might reach a high ratio of the success rates of two groups of users, since the exposed group of users have high purchase intention. If one fairly compares the success rates of the exposed users with the other teenage females with no ad exposure, one can obtain the **true amplifier**, which may be much smaller. Such a fair comparison may not always be available since in observational dataset, the unexposed group is usually the majority of the population, which can be rather different than the exposed group of users. We describe such way of campaign design, namely, targeting the users with high intention to make success actions even without ad exposures and hence showing untruthful high amplifier, ‘smart cheating’.

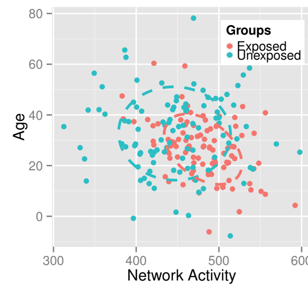

Figure 1: Exposed and Unexposed Groups: Different Network Activity and Age Distributions

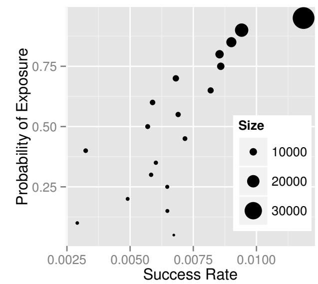

Figure 2: Probability to Be Exposed along with Success Rates

Smart cheating exists extensively in online advertising via targeting the users who are more likely to perform success actions. Even the campaigns without explicitly targeting criterion may also conduct implicit smart cheating. The reason is that, the adversing strategy, especially in real-time bidding (RTB), usually involves success rate prediction that incorporates user characteristics [1]. Hence it may implicitly choose the users who are more likely to perform success actions regardless of ad exposures. By recognizing smart cheating in the campaigns, advertisers can compare the true impact of targeting strategies, and hence choose the campaigns that lead to more incremental success actions. A scientific way to define and quantify smart cheating has long been demanded by the advertisers, and yet few studies have been focused on this topic.

In this paper, we define and quantify smart cheating by measuring the fake inflation of the naive amplifier, comparing to the true amplifier on the exposed users estimated through causal inference, which leads to a smart cheating ratio (SCR). We apply our approach to multiple campaigns, and successfully reveal that some of the campaigns are conducting severe smart cheating.

2. SMART CHEATING

We quantify smart cheating based on the difference of the naive amplifier and the true amplifier. The naive amplifier is formally

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05. http://dx.doi.org/10.1145/2740908.2742758.

defined as follows. In an observational dataset, suppose we observe N users indexed by $i = 1, 2, ..., N$, each with a characteristic vector with K elements $X_i = (X_{i,1}, ..., X_{i,K})$. A user is either exposed to ads or not, and the exposure indicator is represented by z_i, where $z_i = 1$ indicates exposures and $z_i = 0$ indicates no exposure. Hence the users are divided into two groups: the exposed group and control (i.e., unexposed) group. The user-level outcomes, for example purchases, are indicated by y_i, where $y_i = 1$ indicates success action(s) and $y_i = 0$ indicates no success action. The naive success rates of the exposed and control groups are $r_{naive,exposed} = \frac{1}{\sum_i z_i} \sum_i z_i y_i$ and $r_{naive,control} = \frac{1}{\sum_i (1-z_i)} \sum_i (1 - z_i) y_i$ receptively. The naive amplifier is simply the ratio $R_{\mathrm{naive}} = r_{\mathrm{naive},exposed}/r_{\mathrm{naive},control}$.

However, as illustrated in Section 1, the naive amplifier is biased since the control group may have different user characteristics than the exposed group. One would need to 'match' the control group to the exposed group, and then by comparing the observed exposed group and the matched control group, one can obtain the true amplifier on the exposed users. There are various approaches to obtain the success rate of the 'matched' control group [2, 4] through causal inference, utilizing user characteristics X_i for the matching. In this paper, we estimate the success rate of the 'matched' control group with the method described in [3] as $r_{\mathrm{matched},control}$, and obtain the matched success rate ratio R_{matched} as $R_{\mathrm{matched}} = r_{\mathrm{naive},exposed}/r_{\mathrm{matched},control}$. R_{matched} represents the measurement of the true amplifier as it matches the control group to the exposed group.

We quantify smart cheating as the inflation of the naive amplifier R_{naive}, comparing to the true amplifier estimated by R_{matched}. To eliminate the impact of the scale of the amplifier, we define a smart cheating ratio (SCR) to quantify the severity of smart cheating as $SCR = \frac{R_{\mathrm{naive}} - R_{\mathrm{matched}}}{R_{\mathrm{matched}}}$.

The sign of SCR reflects the existence of smart cheating. Specifically we summarize three possible cases as below. 1) $SCR > 0$: The naive amplifier is larger than the truth. This means that the naive amplifier is untruthfully inflated, resulting from the different user characteristics of the exposed and control groups. The larger the absolute value of SCR, the more severe the smart cheating. 2) $SCR < 0$: The naive amplifier is smaller than the truth. The true amplifier is underestimated by naive method, which means the campaign is reaching out to 'hard users', who are not likely to perform success actions comparing to the control group. 3) $SCR = 0$: The naive amplifier equals to the truth. This means that the ads reach users with the same success intention as the control group.

We further embed the SCR calculation in a bootstrap framework to obtain the estimation error and confidence interval (CI), which enables hypothesis testing. We repeatedly generate bootstrap samples (i.e., a set of random samples drawn with replacement from the dataset), and estimate the SCR based on the samples. Suppose we draw B bootstrap samples and the estimated SCR from the B bootstrap samples are $\widehat{SCR}_b^*$, $b = 1, 2, ..., B$ respectively. Based on $\widehat{SCR}_b^*$, one can obtain: 1) the final estimated $\widehat{SCR}$ by averaging $\widehat{SCR}_b^*$; 2) the standard error of the estimated SCR, $\hat{\sigma}_{SCR}$, by calculating the standard deviation of $\widehat{SCR}_b^*$; and 3) the $1 - \alpha$ CI $(q_{\alpha/2}, q_{1-\alpha/2})$, where q_a is the a'th empirical quantile of $\widehat{SCR}_b^*$'s (i.e., q_a is the value such that the proportion of $\widehat{SCR}_b^*$'s smaller than q_a equals to a.). Usually one can set $\alpha = 0.05$, and then the 95% CI is $(q_{0.025}, q_{0.975})$. One can also perform hypothesis test $H_0 : SCR = 0$ Versus $H_a : SCR \neq 0$. One rejects the null hypothesis when 0 is not within the $1 - \alpha$ CI $(q_{\alpha/2}, q_{1-\alpha/2})$.

3. ASSESSMENT OF REAL CAMPAIGNS

Smart cheating exists extensively in online advertising. In this section, we illustrate the smart cheating assessment with multiple real campaigns from a premium Internet media company, involving millions of unique users. We collect 1000 user-level characteristics, including website visitation, ad exposure, demographic information, market interest, etc., repeatedly on a daily basis [1]. We randomly select four campaigns, including companies from wireless service, finance, information technology, and phone system industries, each involving millions of users. The results are shown in Table 1. The SCR's of the first three campaigns are hugely positive under 0.05 significance level and hence indicate severe smart cheating. Note that after matching the control group to the exposed group and hence discounting for smart cheating, the finance industry ad shows negative amplifier, which reveals the negative impact of this campaign. In such case, failure to consider smart cheating may even harm the advertiser. The SCR of the phone system company is negative, meaning that this campaign reaches out to 'hard users' and had positive impact on those users.

Table 1: Compaign Summaries

Advertiser	R_{naive}	$R_{matched}$	$\widehat{SCR}$	$\hat{\sigma}_{SCR}$	CI
Wireless Service	2.52	1.75	0.43	0.03	(0.37, 0.49)
Finance	1.31	0.88	0.33	0.04	(0.25, 0.41)
Information Technology	1.80	1.21	0.33	0.03	(0.27, 0.40)
Phone System	0.51	1.27	-0.60	0.04	(-0.66, -0.53)

To further visualize the targeting bias of the exposed group, we empirically calculate the success probabilities of the exposed group along with ad exposure probabilities. Specifically, we estimate the probability that each user is exposed as $\hat{p}_i = P(z_i = 1|X_i), \forall i$, utilizing the method described in [3]. We determine a series of probability buckets, for example $[0, 0.05), [0.05, 0.10), ..., [0.95, 1]$. For each bucket k, we select the exposed users with $\hat{p}_i$ within the probability bucket, count the number of such exposed users as b_k^e and the number of such exposed users with success action as a_k^e, and calculate the success rate of the corresponding bucket as a_k^e/b_k^e. We then draw the estimated success rates along with the probability buckets. The visualization of the wireless service campaign is demonstrated in Figure 2. The figure shows that, a user with larger success tendency (as in the x-axis) is more likely to be exposed, i.e. the campaign is conducting smart cheating, which confirms the conclusion from SCR.

The results from real-world ad campaigns show that smart cheating exists extensively and can be rather severe in online advertising.

4. REFERENCES

[1] Y. Chen, P. Berkhin, B. Anderson, and N. R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th KDD*, pages 1307–1315. ACM, 2011.

[2] W. Sun, P. Wang, D. Yin, J. Yang, and Y. Chang. Causal inference via sparse additive models with application to online advertising. In *Proceedings of the 29th AAAI*. AAAI, 2015.

[3] P. Wang, Y. Liu, M. Meytlis, H.-Y. Tsao, J. Yang, and P. Huang. An efficient framework for online advertising effectiveness measurement and comparison. In *Proceedings of the 7th WSDM*. ACM.

[4] P. Wang, W. Sun, D. Yin, J. Yang, and Y. Chang. Robust tree-based causal inference for complex ad effectiveness analysis. In *Proceedings of the 8th WSDM*. ACM, 2015.

[1]The reported datasets and results are deliberately incomplete and subject to anonymization, and thus do not necessarily reflect the real portfolio at any particular time.