# Lift-Based Bidding in Ad Selection

Jian Xu
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
xujian@yahoo-inc.com

Xuhui Shao
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
xshao@yahoo-inc.com

Jianjie Ma
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
jianma@yahoo-inc.com

Kuang-Chih Lee
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
kclee@yahoo-inc.com

Hang Qi
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
hangqi@yahoo-inc.com

Quan Lu
Yahoo Inc.
701 First Ave
Sunnyvale, CA 94089
qlu@yahoo-inc.com

## ABSTRACT

Real-time bidding (RTB) has become one of the largest online advertising markets in the world. Today the bid price per ad impression is typically decided by the expected value of how it can lead to a desired action event (e.g., registering an account or placing a purchase order) to the advertiser. However, this industry standard approach to decide the bid price does not consider the actual effect of the ad shown to the user, which should be measured based on the performance lift among users who have been or have not been exposed to a certain treatment of ads. In this paper, we propose a new bidding strategy and prove that if the bid price is decided based on the performance lift rather than absolute performance value, advertisers can actually gain more action events. We describe the modeling methodology to predict the performance lift and demonstrate the actual performance gain through blind A/B test with real ad campaigns in an industry-leading Demand-Side Platform (DSP). We also discuss the relationship between attribution models and bidding strategies. We prove that, to move the DSPs to bid based on performance lift, they should be rewarded according to the relative performance lift they contribute.

## Keywords

Bidding Strategy, Lift Prediction, Attribution, Real-Time Bidding, Online Advertising

## 1. INTRODUCTION

Online advertising is one of the fastest growing industries with $58 billion total spend projected in 2015 in US alone. One of the most significant trends in online advertising in recent years is real-time bidding (RTB), or sometimes more broadly referred to as programmatic buying. In RTB, advertisers have the ability of making decisions whether and how much to bid for every impression that would lead to the best expected outcome (action). It is analogous to stock exchanges in that data-driven algorithms are used to automatically buy and sell ads in real-time. The bidding algorithm can use the contextual and user behavior data to select the best ads, in order to optimize the effectiveness of online advertising.

Demand-Side Platforms (DSPs) are thus created to help advertisers manage their campaigns and optimize their real-time bidding activities. Most advertisers leverage these platforms to optimize the bidding of their ad campaigns. The DSPs offer different pricing models. If the goal of an advertising campaign is reaching a specific audience then Cost Per Mille (CPM) is normally used. However, as most advertisers care more about the how the campaigns lead users to take desired actions such as signing up a service or making a purchase, they would prefer performance based pricing models such as Cost Per Click (CPC) and Cost Per Action (CPA). In addition, there are also hybrid pricing models where the ads are priced by CPM and yet the advertisers have an implied effective Cost Per Click (eCPC) or effective Cost Per Action (eCPA) goal to gauge the optimization results. This paper focuses on the CPA pricing model as it is the most challenging prediction problem and yet the most versatile framework for meeting advertiser's marketing objectives.

State-of-the-art DSPs that support such CPA pricing model typically convert an advertiser's CPA bid to an estimated Cost Per Mille (eCPM) bid in order to participate in the RTB auctions where the winning ad is chosen based on the highest bid. In pure second price auctions, theoretically the optimal bidding strategy is truth-telling: to bid their private values. Therefore, the prevalent practice to derive such eCPM bid is estimating the Action Rate (AR) which is the probability that the impression will lead to a desired action and multiplying it by the CPA bid (i.e., eCPM=AR×CPA). There has been significant work in the literature on estimating AR accurately and also on altering the final bid price to accommodate additional constraints such as budget and smooth delivery. However, there have been very little discussion on whether such prevalent bidding strategy truly benefits the advertisers in terms of bringing the maximal number of actions to the advertisers.

Such bidding strategy neglects the probability that a user

will take the desired action even if the impression is not shown. For example, a loyal Pampers customer will make further purchases even if not exposed to any Pampers ad. An analogy can be found in the political marketing. As early as in 2012, Obama's presidential campaign already focused on the swing voters and used a *persuasion score* to quantify how easily they can be shifted to vote for the Democrat [11]. It is surprising that the online advertising market has lagged behind at this point. We argue that the current prevalent bidding strategy is categorically sub-optimal by design. A blind (not known to the bidding strategy) A/B test can demonstrate this clearly and quantitatively. The effective increased cost to advertisers due to wasted ad impressions can worth billions of dollars each year.

## 1.1  Motivating Examples

The above bidding strategy calculates the bid price based on the AR (proxy for *value* to the advertiser) of a user. The problem with this strategy is that campaign budget would be spent on users who already have high ARs instead of those that could have been greatly influenced by the ads (i.e., those with high AR *lift* because of the ads). The difference is best illustrated by the examples below.

EXAMPLE 1 (VALUE-BASED BIDDING). Suppose a DSP is bidding on behalf of an advertiser to acquire impressions and the advertiser's CPA=$100. Suppose there are two ad requests from user $a$ and $b$ respectively. The AR of user $a$ is 0.04 if he/she is shown the advertiser's ad, otherwise the AR is 0.03. User $b$ has an AR of 0.02 if he/she is shown the ad, otherwise the AR is 0.001.

Based on the common practice in the industry, the bid prices should be the absolute ARs assuming ads are shown times CPA. So the the bid prices for $a$ and $b$ are $0.04 \times \$100 = \$4$ and $0.02 \times \$100 = \$2$ respectively. Suppose the highest bid prices from other competitors are equally $3.5 for both $a$ and $b$, which means the advertiser will win the auction for $a$ while lose the auction for $b$. In this case, the expected total number of actions the advertiser can have is $0.04 + 0.001 = 0.041$, the inventory cost of the DSP is $3.5. Since advertiser only pays for impressions that lead to actions, the expected revenue of the DSP is $4. ∎

EXAMPLE 2 (LIFT-BASED BIDDING). Let us continue with Example 1. It is not difficult to see that user $b$ should be more preferable to the advertiser because the advertiser could expect a more significant AR lift from it (the AR lift of $a$ is $0.04 - 0.03 = 0.01$ while $b$ is $0.02 - 0.001 = 0.019$).

If the bid prices the DSP places on $a$ and $b$ are proportional to the AR lifts, for instance $2 and $3.8, then the advertiser will win the auction for $b$ instead of $a$. In this case, the expected total number of actions the advertiser can have is $0.03 + 0.02 = 0.05$ - better than that in Example 1. The inventory cost of the DSP is also $3.5, but the expected revenue of the DSP becomes $2 because the advertiser only pays for impressions that lead to actions. It results in a negative profit to the DSP. ∎

The above toy examples show that the advertiser eventually gets more actions when bid based on AR lift than based on absolute ARs. In addition, they also show that the advertiser's marketing objectives are not aligned with the DSP's interests. We believe by directly paying for performance lift,

advertisers will get the most benefit and move DSPs to the more optimized bidding behavior - from *value-based bidding* to *lift-based bidding*.

## 1.2  Challenges and Our Contribution

In this paper, we advocate for an industry-wide transition from value-based bidding to lift-based bidding. That is, the bid price should be based on the AR lift instead of the absolute AR. However, it also poses several interesting challenges.

First, existing AR prediction models estimate the absolute ARs assuming the ads are shown to the users. There is little effort in the literature to explicitly model AR lift. Since a user is either shown or not shown an ad, it is theoretically impossible to collect data that cover both cases of an ad request. Therefore, learning from past ad serving logs to estimate the AR lift is a challenging task. Second, although the motivating examples have shown that the prevalent value-based bidding is sub-optimal in terms of bringing maximal number of actions to advertisers, can we prove that lift-based bidding is, if not optimal, a better alternative? Moreover, even if lift-based bidding is a better strategy than value-based bidding to the advertisers, would DSPs adopt such a bidding strategy? As we have shown in Example 2, a DSP may lose money with this strategy.

In this paper, we strive to address these unique and interesting challenges. Our contributions can be summarized as follows:

- We propose the concept of lift-based bidding, which we prove both mathematically and empirically to be a better bidding strategy than value-based bidding in terms of maximizing advertiser benefits.

- We describe a simple yet effective modeling methodology to predict AR lift. Online blind A/B test with real ad campaigns in an industry-leading DSP prove our concepts when we push lift-based bidding to the marketplace.

- We point out that the advertiser's marketing objectives are not aligned with the DSP's interest under the industry standard attribution model. To move the DSPs to lift-based bidding, they should be rewarded according to the relative performance lift they contribute.

## 2.  VALUE-BASED BIDDING VS. LIFT-BASED BIDDING

In this section, we prove that lift-based bidding is a better strategy than value-based bidding in terms of maximizing advertiser benefits. However, it would be opposed by DSPs under the industry-standard last-touch attribution model.

DEFINITION 1 (AR, BACKGROUND AR, AND AR LIFT). *Given an ad request $q$ from a user $u$ and an advertiser $A$, we define AR w.r.t. $(q, u, A)$ as the probability that $u$ will take the desired action defined by $A$ after the ad of $A$ is served to $q$, background AR w.r.t. $(q, u, A)$ as the probability that $u$ will take the desired action if the ad of $A$ is not served to $q$, and AR lift as the difference between AR and background AR. We denote by $p$ the AR, $\Delta p$ the AR lift, and $p - \Delta p$ the background AR if no further specification is made.*

The common practice in the industry is to bid CPA×p. We generalize this practice and define value-based bidding as follows:

DEFINITION 2 (VALUE-BASED BIDDING). *Let $p$ be the AR of a user if the advertiser's ad is shown, value-based bidding places a bid price of $\alpha \times p$ to acquire an impression from this user for the advertiser, where $\alpha > 0$.*

However, examples in Section 1.1 show that such bidding strategy does not necessarily optimize the overall campaign performance for the advertiser. As we advocate for focusing more on the AR lift instead of absolute AR, we propose the concept of lift-based bidding in which the bid price is proportional to the AR lift.

DEFINITION 3 (LIFT-BASED BIDDING). *Let $\Delta p$ be the AR lift of a user if the advertiser's ad is shown, lift-based bidding places a bid price of $\beta \times \Delta p$ to acquire an impression from this user for the advertiser, where $\beta > 0$.*

In the CPA pricing model, a DSP is rewarded based on the number of actions attributed to it. Advertisers finally pay DSPs based on the amount of actions attributed to them. The industry standard attribution model is *last-touch attribution*.

DEFINITION 4 (LAST-TOUCH ATTRIBUTION). *An advertiser attributes the full credit of an observed action to the last relevant ad impression and hence the DSP which delivered that impression.*

Suppose there are two DSPs $DSP_1$ and $DSP_2$ bidding on behalf of the same advertiser at the same time. $DSP_1$ practices value-based bidding while $DSP_2$ executes lift-based bidding. Let $u_i$ be user of the $i$-th ad request, $p_i$ be the AR if the advertiser's ad is shown to $u_i$, and $\Delta p_i$ be the AR lift because of the ad impression. To simplify the discussion, let us assume every ad request is from a different user and there are no additional candidates in the auctions. In an ad exchange marketplace with pure second-price auction, we have

LEMMA 1. $DSP_1$ *wins the auction for $u_i$ at the cost of $\beta \times \Delta p_i$ if $\alpha \times p_i > \beta \times \Delta p_i$; $DSP_2$ wins the auction for $u_i$ at the cost of $\alpha \times p_i$ if $\alpha \times p_i < \beta \times \Delta p_i$.*

Lemma 1 points out that $DSP_1$'s cost when winning is actually $DSP_2$'s bid due to second-price auction, and vise-versa.

THEOREM 1. *With the last-touch attribution model, $DSP_2$ yields more actions than $DSP_1$ for the advertiser when the advertiser attributes the same amount of actions to them* [1].

*Proof.* Let $i$ be the index of all the users, $j$ be the index of those users that $DSP_1$ wins (i.e., $\alpha \times p_j > \beta \times \Delta p_j$), and $k$ be

---
[1] The condition of $DSP_1$ and $DSP_2$ getting equal attributions from the advertiser is an important setup to illustrate that DSP2 will in fact produce relatively more actions for the advertiser. This setup condition can be achieved by adjusting the winning landscape through the parameters of $\alpha$ and $\beta$

the index of those users that $DSP_2$ wins (i.e., $\alpha \times p_k < \beta \times \Delta p_k$). It is straight forward to see that the expected number of actions to be attributed to $DSP_1$ and $DSP_2$ are $\sum_j p_j$ and $\sum_k p_k$ respectively. The expected number of actions if only $DSP_1$ is considered can be decomposed as two parts: sum of the ARs of users that $DSP_1$ wins, and sum of the background ARs of users that $DSP_1$ loses. So it becomes $\sum_j p_j + \sum_k (p_k - \Delta p_k)$. Similarly, the expected number of actions if only $DSP_2$ is considered is $\sum_j (p_j - \Delta p_j) + \sum_k p_k$.

Therefore, let $\mathscr{A}_1$ ($\mathscr{A}_2$) be the expected number of actions per attributed action if only $DSP_1$ ($DSP_2$) is considered, we have

$$\mathscr{A}_1 = \frac{\sum_j p_j + \sum_k (p_k - \Delta p_k)}{\sum_j p_j} \qquad (1)$$

$$\mathscr{A}_2 = \frac{\sum_j (p_j - \Delta p_j) + \sum_k p_k}{\sum_k p_k} \qquad (2)$$

When the same amount of actions is attributed to $DSP_1$ and $DSP_2$ (i.e., $\sum_j p_j = \sum_k p_k$), by swapping the denominators and consolidating the numerators in Equation 1 and 2, noticing $\alpha \times p_j > \beta \times \Delta p_j$ and $\alpha \times p_k < \beta \times \Delta p_k$, we have

$$\mathscr{A}_1 = \frac{\sum_i p_i - \sum_k \Delta p_k}{\sum_k p_k} < \frac{\sum_i p_i}{\sum_k p_k} - \frac{\alpha}{\beta} \qquad (3)$$

$$\mathscr{A}_2 = \frac{\sum_i p_i - \sum_j \Delta p_j}{\sum_j p_j} > \frac{\sum_i p_i}{\sum_j p_j} - \frac{\alpha}{\beta} \qquad (4)$$

Therefore it is obvious that $\mathscr{A}_1 < \mathscr{A}_2$. ∎

THEOREM 2. *With the last-touch attribution model, $DSP_2$ costs more than $DSP_1$ when the advertiser attributes the same amount of actions to them.*

*Proof.* Again, let $i$ be the index of all the users, $j$ be the user index such that $\alpha \times p_j > \beta \times \Delta p_j$, and $k$ be the user index such that $\alpha \times p_k < \beta \times \Delta p_k$. The expected number of actions to be attributed to $DSP_1$ and $DSP_2$ are $\sum_j p_j$ and $\sum_k p_k$ respectively. The cost of $DSP_1$ and $DSP_2$ are $\sum_j \beta \times \Delta p_j$ and $\sum_k \alpha \times p_k$ respectively. Therefore let $\mathscr{C}_1$ and $\mathscr{C}_2$ be the cost per attributed action of $DSP_1$ and $DSP_2$ respectively, we have

$$\mathscr{C}_1 = \frac{\sum_j \beta \times \Delta p_j}{\sum_j p_j} \qquad (5)$$

$$\mathscr{C}_2 = \frac{\sum_k \alpha \times p_k}{\sum_k p_k} \qquad (6)$$

Noticing $\alpha \times p_j > \beta \times \Delta p_j$, we have $\mathscr{C}_1 < \beta \times \frac{\alpha}{\beta} = \alpha$. It is apparent that $\mathscr{C}_2 = \alpha$, therefore $\mathscr{C}_1 < \mathscr{C}_2$. ∎

We have shown that lift-based bidding benefits the advertisers but the resulting higher cost per attributed action may undermine the interests of the DSPs. The root cause of this conflict lies in the fact that the industry-standard last-touch attribution model does not attribute actions fairly to the DSPs. Researchers have pointed out that an action should be attributed to multiple touch points in a data driven fashion [12] and/or based on causal lift [5]. In Section 5, we discuss the relationship between attribution models and bidding strategies. We show that to motivate the DSPs to

move from value-based bidding to lift-based bidding, they should be attributed based on the relative action lift they contributed to the final actions.

# 3. LIFT-BASED BIDDING IN ACTION

## 3.1 Predicting AR Lift

Existing user action modeling tries to estimate the absolute AR of a user if an ad is shown. However, to implement the lift-based bidding strategy, it is more important to estimate the AR lift. One may think of building a machine learning model to predict the lift directly. However, since the real ad serving logs contain only instances that an ad is either shown or not shown, it is theoretically impossible to have the true AR lift data for modeling. Therefore, in order to predict the AR lift, we strive to estimate both the ARs assuming the ad is shown or not shown respectively.

Formally, let $a$ be an ad, $s$ be the state of a user at ad request time, and $s_+(a)$ be the state of the user if $a$ is shown. Conceptually, $s$ consists of the user's demographic status, timestamped past events including page views, searches, ad views/clicks, and anything you can think of to describe the user state at ad request time. The only difference between $s$ and $s_+(a)$ is the ad impression of $a$. Let $p(action|s)$ be the AR of the user if $a$ is not shown and $p(action|s_+(a))$ be the AR if $a$ is shown, the AR lift is therefore

$$\Delta p = p(action|s_+(a)) - p(action|s) \qquad (7)$$

Considering a specific ad request instance (i.e., a specific user with a specific state) in the ad serving log, the ad was either shown or not shown. Therefore, either case will be absent in the modeling data. We address this challenge by establishing a model that has sufficient generalization capability. More specifically, we use a function $F$ to map a state to a set of features that are shared among different instances. Then a single and generic AR prediction model $\hat{P}$ is built upon the derived feature set and the AR lift can be estimated as

$$\widehat{\Delta p} = \hat{P}(action|F(s_+(a))) - \hat{P}(action|F(s)) \qquad (8)$$

where the difference between $F(s_+(a))$ and $F(s)$ is reflected by different feature values induced from $a$. At ad serving time when $\Delta p$ is to be estimated for ad $a$, if for instance we consider impression frequency of $a$ as a feature, the feature value in $F(s_+(a))$ should be greater than that in $F(s)$ by one.

## 3.2 Model Training

Our task is to train a generic AR prediction model $\hat{P}$ which can give AR estimations for both cases when an ad is shown or not shown, i.e., $p(action|s_+(a))$ and $p(action|s)$. Existing AR prediction models in the literature are trained based on the post-view [7] or post-click [10, 3] actions. That is, the training samples are collected from only those impression events or click events. For example, in post-view action modeling, each impression event will trigger a training sample which is labeled as positive if there is an action followed. However, such methodology is not preferred in our scenario for several reasons.

First, since the training samples from only impression or click events are not representative to those cases when the ad is not shown, models trained upon these samples are not
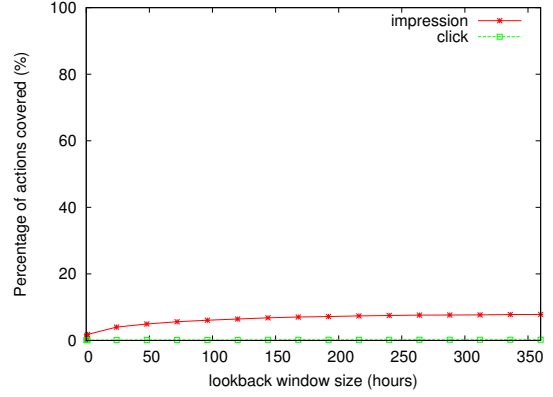


Figure 1: Less than 10% of the reported actions have precedent ad impressions from the same advertiser within the lookback window. Training samples generated only from impression and/or click events miss a large portion of the informative actions.

generalized enough to predict $p(action|s)$. Second, even for predicting $p(action|s_+(a))$, leveraging training samples from only impression or click events still suffers from survival bias. In the RTB marketplace, impressions are purchased through public auctions. Therefore, impressions and clicks are available from only those winning auctions. Such survival bias is prevalent in click modeling. In action modeling, it could have been avoided because actions can happen even there was no impression showed. Third, our observation from real ad campaigns shows that usually for the majority of actions, we have not shown any ad of the advertiser before[2]. In other words, if training samples are only generated from impression or click events, the majority of actions (positive samples) are not leveraged for modeling. Figure 1 shows a typical advertiser whose impressions and clicks only cover less than 10% of actions.

Given these considerations, we take a different approach that we train AR prediction models upon the whole population. Training samples are generated from every user's timeline instead of from merely impression/click events. To mimic the true action distribution, we first randomly select a user $u$ weighted by its ad request frequency (i.e., a heavy user is more likely to be selected). Then a random timestamp $ts$ is chosen on $u$'s timeline and a training sample is generated based on $u$ and $ts$. If $u$ has at least one action in the *action window* (denoted by $(ts, ts + aw]$), the sample is labeled as positive and otherwise negative, where $aw$ is the action window size typically several hours to a couple of days according to different business definitions. Then features are generated within the *feature window* (denoted by $(ts - fw, ts]$), where $fw$ is the feature window size.

Raw input variables for feature generation include user historical profile within the feature window with such details as page views, ad impressions, clicks, searches and mobile app based events. Each user event is used as a point-in-time to generate features. Example includes number of ad impressions in the last hour by a given advertiser. Table 5

---

[2]Advertisers usually report actions via action pixels placed on their websites or apps. Therefore DSPs can have the full action set.

is a list of different types of features we generated for AR prediction. At serving time, ad request details such as geo-location, web page or mobile app being visited are folded in these features so that the run-time context can also be leveraged for prediction. For example, if the recency of visiting Yahoo! homepage is a feature, an ad request from Yahoo! homepage will set this feature value as "most recent".

The sample generation terminates whenever all the action events have been involved or the positive samples are sufficient. Once the training samples are gathered, we train a Gradient-Boosting-Decision-Tree (GBDT) model to predict the rank order and then calibrate using isotonic regression to translate a GBDT score to an AR. Please note that we utilize our in-house GBDT tool with distributed training capability for modeling; however, it is only one alternative and any machine learning algorithm can be applied.

## 3.3 Fitting Lift-based Bidding in the Marketplace

Conventional value-based bidding typically calculate the bid price by multiplying predicted absolute AR by advertiser CPA. In lift-based bidding, it is apparently not proper to simply multiply AR lift by the same advertiser CPA. Otherwise one can seldom win any auction if the majority of the other competitor DSPs are still practicing value-based bidding. Recall that within lift-based bidding the bid price is proportional to the AR lift i.e., $\beta \times \Delta p$. Our selection of $\beta$ is therefore

$$\beta = \frac{\bar{p}}{\overline{\Delta p}} \times \text{CPA} \tag{9}$$

where $\bar{p}$ is the population mean of the AR and $\overline{\Delta p}$ is the population mean of AR lift. The idea is straightforward: if the advertiser is willing to pay CPA for each action in the conventional way, then each incremental action should be paid at the price of $\frac{\bar{p}}{\overline{\Delta p}} \times$CPA if only incremental actions need to get paid.

## 4. BLIND A/B TEST WITH A REAL DSP

To empirically prove our proposed concepts, we set up a blind A/B test with a real DSP. We conducted our experiments on Yahoo Ad Manager Plus (YAM+), one of the largest DSPs in the world. The blind A/B test setup means that users are randomly split into distinct groups not known to the bidding models. More specifically, we first randomly split users into three equal-sized buckets. Then we created three bidders: a "no bid" bidder which always places a zero bid, a value-based bidder, and a lift-based bidder. We selected five advertisers to participate in the test. To be fair, each advertiser's budget was evenly split and assigned to the value-based bidder and lift-based bidder respectively (we note that the "no bid" bidder would not spend any budget). Each advertiser's campaign ran for one week and their budgets were all spent out which means value-based bidder and lift-based bidder spent the same amount of budget. We counted the number of actions[3] observed in each bucket in a three weeks window from the campaign start date. The result shown in Table 1, 2, 3 and 4 backed up our claims

---

[3]Action pixel fires to be more accurate. Even in the "no bid" bucket, there can still be actions reported via action pixel fires on advertiser's website or app.

and methods.

First, by comparing result from "no bid" bidder and value-based bidder (Table 1), as we mentioned in Section 1, it is easy to demonstrate and quantify that a user will take the desired action even if the impression is not shown. Although showing ads to a substantial number of users did help lift the action yield, the number of actions even if no ad impression is shown (or in other words background AR) is already significant. This is not surprising because advertisers typically run their campaigns through multiple channels such as TV, magazine, internet, etc and in multiple formats such as search, display, video etc simultaneously. Even if a user is not shown any display ad by our DSP, he/she can still be influenced by a variety of other touch points. This is exactly why lift-based bidding is more preferable since it tries to maximize the effectiveness of display ads by taking background AR into account. Table 2 shows the comparison between "no bid" bidder and lift-based bidder.

Second, from advertiser's perspective, lift-based bidding generated more actions than value-based bidding with the same amount of budget. This result is observed consistently among all the five pilot advertisers (Table 3). Since we have shown that the background actions prevalently exist, an even fairer comparison between lift-based bidding and value-based bidding should be comparing their action lifts over background actions, which we call *lift-over-lift*. Take Advertiser 1 for example, value-based bidder generated 11.2% more actions than the background actions while lift-based bidder yields 28.7% more actions than the background actions. In this case the lift-over-lift measure is $(28.7 - 11.2)/11.2 = 156\%$. Lift-based bidding dramatically increases the *incremental* actions compared to value-based bidding so the lift-over-lift measure can be as high as more than 100% for three of the advertisers.

Third, from DSP's perspective, lift-based bidding resulted in more inventory cost than value-based bidding when the same number of actions are attributed to them. Recall that advertiser only pays for each attributed action at the price of CPA and the two bidders spent the same amount of budget, therefore the number of attributions are the same. From Table 4 we observe the inventory cost of lift-based bidder is consistently higher than that of value-based bidder among the 4 campaigns.

Lastly and interestingly, we have observed increased number of impressions when comparing lift-based bidding to value-based bidding. Even though the overall inventory cost is higher, the effective cost of bidding is lower for lift-based bidding than value-based bidding. The lift-based bidder does not always compete with other bidders for those high AR users. Instead, it tries to acquire the users who are more likely to be influenced. Therefore it has the advantage of avoiding competition and acquiring more impressions at a lower cost-per-impression.

The above blind A/B test result with real online ad campaigns backed up our concepts and techniques. Since the lift-based bidder took the risk of losing revenue due to the current attribution model while as we have seen it actually benefits the advertisers, advertisers should think of a more reasonable attribution model to align the DSP's benefits with their marketing objectives. In Section 5, we discuss the relationship between attribution and bidding.

## 5. ATTRIBUTION AND BIDDING

| Adv | No bid | | Value-based bidding | | Incremental action | Action lift |
|---|---|---|---|---|---|---|
| | # imps | # actions | # imps | # actions | | |
| 1 | 0 | 642 | 53,396 | 714 | 72 | **11.2%** |
| 2 | 0 | 823 | 298,333 | 896 | 73 | **8.9%** |
| 3 | 0 | 1,438 | 11,048,583 | 1,477 | 39 | **2.7%** |
| 4 | 0 | 1892 | 3,915,792 | 2,016 | 124 | **6.6%** |
| 5 | 0 | 5,610 | 6,015,322 | 6,708 | 1,098 | **19.6%** |

**Table 1: Blind A/B test result on five pilot advertisers - Value-based bidding v.s. "No bid".**

| Adv | No bid | | Lift-based bidding | | Incremental action | Action lift |
|---|---|---|---|---|---|---|
| | # imps | # actions | # imps | # actions | | |
| 1 | 0 | 642 | 59,703 | 826 | 184 | **28.7%** |
| 2 | 0 | 823 | 431,637 | 980 | 157 | **19.1%** |
| 3 | 0 | 1,438 | 11,483,360 | 1509 | 71 | **4.9%** |
| 4 | 0 | 1892 | 4,368,441 | 2,471 | 579 | **30.6%** |
| 5 | 0 | 5,610 | 8,770,935 | 8,291 | 2,681 | **47.8%** |

**Table 2: Blind A/B test result on five pilot advertisers - Lift-based bidding v.s. "No bid".**

| Adv | Value-based bidding | | | Lift-based bidding | | | Action lift | Lift-over-lift |
|---|---|---|---|---|---|---|---|---|
| | # imps | # actions | Action lift (vs "No bid") | # imps | # actions | Action lift (vs "no bid") | | |
| 1 | 53,396 | 714 | 11.2% | 59,703 | 826 | 28.7% | 13.6% | **156%** |
| 2 | 298,333 | 896 | 8.9% | 431,637 | 980 | 19.1% | 9.4% | **115%** |
| 3 | 11,048,583 | 1,477 | 2.7% | 11,483,360 | 1509 | 4.9% | 2.2% | **82%** |
| 4 | 3,915,792 | 2,016 | 6.6% | 4,368,441 | 2,471 | 30.6% | 22.6% | **367%** |
| 5 | 6,015,322 | 6,708 | 19.6% | 8,770,935 | 8,291 | 47.8% | 23.6% | **144%** |

**Table 3: Lift-based bidding v.s. Value-based bidding - Advertiser's perspective. "Action lift" is the absolute # actions difference between lift-based bidding and value-based bidding. "Lift-over-lift" is comparing the their action lifts over "no bid".**

| Adv | Value-based bidding | | | Lift-based bidding | | | Inventory cost diff | Cost per imp diff |
|---|---|---|---|---|---|---|---|---|
| | # imps | # attrs | Inventory cost | # imps | # attrs | Inventory cost | | |
| 1 | 53,396 | 50 | $278.73 | 59,703 | 50 | $300.31 | **7.7%** | **-3.6%** |
| 2 | 298,333 | 80 | $1,065.05 | 431,637 | 80 | $1,467.57 | **37.8%** | **-4.8%** |
| 3 | 11,048,583 | 240 | $25,522.22 | 11,483,360 | 240 | $25,837.56 | **1.2%** | **-2.6%** |
| 4 | 3,915,792 | 200 | $10,846.74 | 4,368,441 | 200 | $11,183.21 | **3.1%** | **-7.6%** |
| 5 | 6,015,322 | 500 | $19,296.51 | 8,770,935 | 500 | $23,501.90 | **21.8%** | **-16.5%** |

**Table 4: Lift-based bidding v.s. Value-based bidding - DSP's perspective. Both bidders spent out equal amount of assigned budget, so the # attributions are always the same. Cost per impression is the inventory cost averaged by # impressions.**

We have proved in Section 2 that DSPs may be opposed to lift-based bidding because of higher cost per attributed action. The root cause is that they are not attributed and hence rewarded based on action lift. Therefore they do not have the incentive to bid to maximize total actions. Actually, a rational DSP will always bid at the price

$$\text{eCPM} = \text{AR} \times \text{CPA} \times p(attribution|action) \quad (10)$$

where $p(attribution|action)$ is the probability it gets attributed if an action happens. The industry common practice to bid eCPM=AR×CPA is simplifying this by assuming that the full credit of an action will be eventually attributed to the DSP. In many scenarios, such assumption is true when the attribution model is last-touch attribution. However, we must point out that such assumption is not always valid especially when there are multiple DSPs running campaigns for the same advertiser simultaneously.

Given that the DSPs will always bid a rational eCPM price, we are more interested in how advertisers can move them to lift-based bidding, which we have shown both theoretically and empirically can bring more actions to the advertisers. Intuitively, if the DSPs are attributed based on the relative AR lift they contribute to the final AR, they have more incentive to practice lift-based bidding. Therefore the key to the solution of the above problem is the attribution model.

Again, let $u_i$ be the user of the $i$-th ad request, $p_i$ be the AR if the advertiser's ad is shown to $u_i$, and $\Delta p_i$ be the AR lift because of the impression. Let $a_i = p(attribution|action, u_i)$ be the probability that the action from $u_i$ is attributed to the DSP that wins $u_i$. In other words, $a_i$ encodes the attribution model. Suppose there are two DSPs $DSP_3$ and $DSP_2$ bidding on behalf of the same advertiser at the same time. $DSP_3$ always bid the rational price (i.e., bid $CPA \times p_i \times a_i$) while $DSP_2$ practices lift-based bidding (i.e., bid $\beta \times \Delta p_i$).

| Source | Feature name | Note |
|---|---|---|
| behaviors | IMP_FREQ_ADV_i IMP_RNCY_ADV_i CLK_FREQ_ADV_i CLK_RNCY_ADV_i PV_FREQ_TOPIC_j PV_RNCY_TOPIC_j SRCH_FREQ_TOPIC_j SRCH_RNCY_TOPIC_j | impression/click frequency and recency from each advertiser, page view/search frequency of each topic (pages and search queries are segmented into several semantic topics) |
| demographics | AGE_BUCKET GENDER GEO_AREA | age is bucketized into several age groups, user's geographic location is at some area resolution. |
| mobile | INST_FREQ_APP_k INST_RNCY_APP_k USE_FREQ_APP_k USE_RNCY_APP_k | installation frequency and recency of each mobile app |

**Table 5: Features generated for AR modeling.**

THEOREM 3. *Unless* $a_i = \frac{\beta \times \Delta p_i}{CPA \times p_i}$ *(i.e., $DSP_3$ always bid the same price as $DSP_2$), $DSP_2$ yields more actions than $DSP_3$ for the advertiser when the advertiser attributes the same amount of actions to them.*

*Proof.* Let $i$ be the index of all the ad requests, $j$ be the index of those requests that $DSP_3$ wins (i.e., $CPA \times p_j \times a_j > \beta \times \Delta p_j$), and $k$ be the index of those requests that $DSP_2$ wins (i.e., $CPA \times p_j \times a_j < \beta \times \Delta p_j$). It is straight forward to see that the expected number of actions to be attributed to $DSP_3$ and $DSP_2$ are $\sum_j p_j \times a_j$ and $\sum_k p_k \times a_k$ respectively. The expected number of actions if only $DSP_3$ is considered is $\sum_j p_j + \sum_k (p_k - \Delta p_k)$, and the expected number of actions if only $DSP_2$ is considered is $\sum_j (p_j - \Delta p_j) + \sum_k p_k$.

Therefore, let $\mathscr{A}_3$ ($\mathscr{A}_2$) be the expected number of actions per attributed action if only $DSP_3$ ($DSP_2$) is considered, we have

$$\mathscr{A}_3 = \frac{\sum_j p_j + \sum_k (p_k - \Delta p_k)}{\sum_j p_j \times a_j} \qquad (11)$$

$$\mathscr{A}_2 = \frac{\sum_j (p_j - \Delta p_j) + \sum_k p_k}{\sum_k p_k \times a_k} \qquad (12)$$

If $DSP_3$ and $DSP_2$ are not always bidding the same price, we can always adjust $\beta$ to control the winning landscape so that $DPS_3$ and $DSP_2$ get the same amount of attribution. When the same amount of actions is attributed to $DSP_3$ and $DSP_2$ (i.e., $\sum_j p_j \times a_j = \sum_k p_k \times a_k$), by swapping the denominators and consolidating the numerators in Equation 1 and 2, noticing $CPA \times p_j \times a_j > \beta \times \Delta p_j$ and $CPA \times p_k \times a_k < \beta \times \Delta p_k$, we have

$$\mathscr{A}_3 = \frac{\sum_i p_i - \sum_k \Delta p_k}{\sum_k p_k \times a_k} < \frac{\sum_i p_i}{\sum_k p_k \times a_k} - \frac{CPA}{\beta} \qquad (13)$$

$$\mathscr{A}_2 = \frac{\sum_i p_i - \sum_j \Delta p_j}{\sum_j p_j \times a_j} > \frac{\sum_i p_i}{\sum_j p_j \times a_j} - \frac{CPA}{\beta} \qquad (14)$$

Therefore it is obvious that $\mathscr{A}_3 < \mathscr{A}_2$ unless $DSP_3$ always bid the same price as $DSP_2$ i.e., $a_i = \frac{\beta \times \Delta p_i}{CPA \times p_i}$. ∎

THEOREM 4. *Unless* $a_i = \frac{\beta \times \Delta p_i}{CPA \times p_i}$ *(i.e., $DSP_3$ always bid the same price as $DSP_2$), $DSP_2$ costs more than $DSP_3$*

*when the advertiser attributes the same amount of actions to them.*

*Proof.* The expected number of actions to be attributed to $DSP_3$ and $DSP_2$ are $\sum_j p_j \times a_j$ and $\sum_k p_k \times a_k$ respectively. The cost of $DSP_3$ and $DSP_2$ are $\sum_j \beta \times \Delta p_j$ and $\sum_k CPA \times p_k \times a_k$ respectively. Therefore let $\mathscr{C}_3$ and $\mathscr{C}_2$ be the cost per attributed action of $DSP_3$ and $DSP_2$ respectively, we have

$$\mathscr{C}_3 = \frac{\sum_j \beta \times \Delta p_j}{\sum_j p_j \times a_j} \qquad (15)$$

$$\mathscr{C}_2 = \frac{\sum_k CPA \times p_k \times a_k}{\sum_k p_k \times a_k} \qquad (16)$$

Noticing $CPA \times p_j \times a_j > \beta \times \Delta p_j$, we have $\mathscr{C}_3 < CPA$. It is apparent that $\mathscr{C}_2 = CPA$, therefore $\mathscr{C}_3 < \mathscr{C}_2$ unless $DSP_3$ always place the same bid as $DSP_2$, i.e., $a_i = \frac{\beta \times \Delta p_i}{CPA \times p_i}$. ∎

The above theorems suggest that for a rational DSP, the only way to move it to lift-based bidding is to attribute based on the relative lift it contributes to the final action. We believe by directly paying for the performance lift, the advertisers can move the DSPs to the more optimized bidding strategy.

## 6. RELATED WORK

Predicting AR for display advertising has attracted much research interest in recent years. Most of the existing works towards this goal model actions and clicks within the same framework[7, 10, 8, 2]. In [3], the author pointed out that an action can happen days or weeks later after an impression and trained an additional delay prediction model jointly with the action prediction model. While this share some common thoughts with our action modeling methodology, we notice that none of these existing works explicitly models AR lift.

Although bidding proportionally to absolute AR (e.g. often AR×CPA) has been a common practice in the industry, researchers are exploring alternatives in order to optimize some Key Performance Indicator (KPI) under budget constraint throughout the campaign's lifetime [14, 9]. However, our work is fundamentally different from these bid optimization works. First, they assume a CPM pricing model and therefore bid optimization is under budget constraint, which is not a serious concern in CPA pricing model. Second, in their approaches, the modified bids are still derived from functions that map absolute AR to the final bid. So they still bid based on the absolute AR instead of AR lift.

Last-touch attribution has been used as the default attribution model in the industry. An alternative to the last-touch attribution is multi-touch attribution which attempts to assign credit to multiple touch points when more than one ad are shown to a user with action. Instead of assigning credit based on some predefined rules, several data-driven multi-touch attribution methods have been proposed [12, 5, 15, 13]. Our focus in this paper is more on illustrating the relationship between bidding strategies and attribution models.

There is earlier body of work that applies MDP and reinforcement learning to marketing optimization where the effect of a marketing action is relativized to the reward that

can already be expected [1]. However, in our problem setting, the action to take is deciding a bid price. It brings more complexity that whether a user can be eventually touched by an ad depends on the auction result. The relationship between bid price and bid win-rate is usually non-trivial and highly dynamic. Therefore it poses the challenge that given any bid price, the user may have a probability not touched by the ad and this probability is hard to estimate.

There are some other recent efforts related to our work. In [4], the authors also pointed out that some marketers may prefer to evaluate and optimize campaigns for *incremental* purchases. However, their focus was identifying better proxies than clicks to evaluate online advertising instead of innovating bidding strategies. Budget allocation based on multi-touch attribution is studied in [6], which we claim can be complementary to our work. After budget allocation is done, our approach can further optimize in the individual impression level.

## 7. CONCLUSION AND FUTURE WORK

We have presented a general modeling framework to predict the performance lift, and apply it to drive a new bidding strategy in the RTB environment. We have proven both mathematically and empirically that the new lift-based bidding strategy indeed helps advertisers to improve the campaign performance. In addition, our current implementation can easily fit into an industry-leading DSP and handle up to a million of ad requests per second and we think it can scale to even more.

We believe it will be an industry trend that advertisers will pay only for those impressions that drive incremental market values. This is an exciting new research area with many potential directions. For example, as we have discussed in this paper, predicting the performance lift accurately is a critical task which deserves much research work. In addition, traditionally DSPs place the bid of $AR \times CPA$ where CPA, as the scaling factor, is determined by advertisers with sufficient market knowledge. However, what value should the scaling factor be in lift-based bidding? We have proposed a simple and empirical answer in Section 3.3 but it remains an interesting topic and challenge to the market. Finally, we have advocated for an attribution model to attribute based on relative performance lift. It would be an interesting topic how such attribution model can be implemented in the industry where there are typically several parties involved including advertisers, DSPs, and third party attributors such as DoubleVerify.

## 8. REFERENCES

[1] N. Abe, N. Verma, C. Apte, and R. Schroko. Cross channel optimized marketing by reinforcement learning. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 767–772. ACM, 2004.

[2] D. Agarwal, R. Agrawal, R. Khanna, and N. Kota. Estimating rates of rare events with multiple hierarchies through scalable log-linear models. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–222. ACM, 2010.

[3] O. Chapelle. Modeling delayed feedback in display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1097–1105. ACM, 2014.

[4] B. Dalessandro, R. Hook, C. Perlich, and F. Provost. Evaluating and optimizing online advertising: Forget the click, but there are good proxies. 2012.

[5] B. Dalessandro, C. Perlich, O. Stitelman, and F. Provost. Causally motivated attribution for online advertising. In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, page 7. ACM, 2012.

[6] S. C. Geyik, A. Saxena, and A. Dasdan. Multi-touch attribution based budget allocation in online advertising. In *Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1–9. ACM, 2014.

[7] K.-C. Lee, B. Orten, A. Dasdan, and W. Li. Estimating conversion rate in display advertising from past performance data. *ACM SIGKDD Conf. on Knowledge Discovery and Data Mining*, 2012.

[8] A. K. Menon, K.-P. Chitrapura, S. Garg, D. Agarwal, and N. Kota. Response prediction using collaborative filtering with hierarchies and side-information. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 141–149. ACM, 2011.

[9] C. Perlich, B. Dalessandro, R. Hook, O. Stitelman, T. Raeder, and F. Provost. Bid optimizing and inventory scoring in targeted online advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 804–812. ACM, 2012.

[10] R. Rosales, H. Cheng, and E. Manavoglu. Post-click conversion modeling and analysis for non-guaranteed delivery display advertising. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 293–302. ACM, 2012.

[11] J. Rutenberg. Data you can believe in. *New York Times Magazine*, pages 22–29, 2013.

[12] X. Shao and L. Li. Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 258–264. ACM, 2011.

[13] D. A. Wooff and J. M. Anderson. Time-weighted multi-touch attribution and channel relevance in the customer journey to online purchase. *Journal of Statistical Theory and Practice*, (ahead-of-print):1–23, 2014.

[14] W. Zhang, S. Yuan, and J. Wang. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1077–1086. ACM, 2014.

[15] Y. Zhang, Y. Wei, and J. Ren. Multi-touch attribution in online advertising with survival theory. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pages 687–696. IEEE, 2014.