Causal Models for Real Time Bidding with Repeated User Interactions

Martin Bompaire, Alexandre Gilotte, Benjamin Heymann

Causal Models for Real Time Bidding with Repeated User Interactions

Martin Bompaire*
Criteo AI Lab
Paris, France
m.bompaire@criteo.com

Alexandre Gilotte*
Criteo AI Lab
Paris, France
a.gilotte@criteo.com

Benjamin Heymann*
Criteo AI Lab
Paris, France
b.heymann@criteo.com

ABSTRACT

A large portion of online advertising displays are sold through an auction mechanism called Real Time Bidding (RTB). Each auction corresponds to a display opportunity, for which the competing advertisers need to precisely estimate the economical value in order to bid accordingly. This estimate is typically taken as the advertiser's payoff for the target event – such as a purchase on the merchant website attributed to this display – times this event estimated probability. However, this greedy approach is too naive when several displays are shown to the same user. The purpose of the present paper is to discuss how such an estimation should be made when a user has already been shown one or more displays. Intuitively, while a user is more likely to make a purchase if the number of displays increases, the marginal effect of each display is expected to be decreasing. In this work, we first frame this bidding problem with repeated user interactions by using causal models to value each display individually. Then, based on this approach, we introduce a simple rule to improve the value estimate. This change shows both interesting qualitative properties that follow our previous intuition as well as quantitative improvements on a public data set and online in a production environment.

KEYWORDS

real-time bidding; causality; incrementality; attribution

1 INTRODUCTION

Display advertising is a form of online advertising in which a marketer pays a website owner (the publisher) for the right to show banners to its visitors (users) in the hope of triggering some future sales. It is a large industry: last year, it was estimated that display advertising generated more than 57 billions USD in the United States [Fisher, 2019].

When the internet user reaches the publisher page, he triggers a complex mechanism called Real Time Bidding (RTB). Real Time Bidding involves several intermediaries (DSP, Ad exchange, SSP, ...) in a chain of calls that finishes its run in the advertisers' server in the form of a bid request. In each advertiser's server lives an algorithmic bidding agent (the *bidder*) that implements the advertiser's strategy. When the bidder receives the bid request, it only has a few milliseconds to answer with a bid. Then, the highest bidder receives the right to show a banner to the user, in exchange for a payment that depends on the chosen auction system. The bidder can make a marketing campaign succeed or fail, because it decides for whom, when, where and at which price to buy a banner.

The bidding strategy heavily relies on the display valuation, usually predicted with a machine learning algorithm. The purpose of the present paper is to discuss how such an estimation should be made when a user has already been shown one or more displays. More precisely we propose to correct the greedy formula that values the action of making a display $D_t = 1$ to a user in a state x_t as the product of the target event payoff CPA (Cost per Action) and the expected number of target events S_t that will be attributed to this display:

$$CPA \cdot \mathbb{E}(S_t | X_t = x_t, D_t = 1).$$

The target event S_t is typically a purchase of the user made on the merchant website after he clicked on this particular display D_t. But when a user has clicked on several displays all the credit will be arbitrary assigned to only one of them, typically the last clicked display before the purchase. As a consequence, a single display in the sequence will be labeled with $S_t = 1$ while all the other will be labeled $S_t = 0$, implying that the machine learning algorithm will learn them as useless. We argue that this arbitrary assignment is sub-optimal and propose to value a display with the expected number of *additional* target events in the future ΔS whether they are attributed to this display D_t or not. To do so, we introduce a new factor $\alpha(x_t)$ that estimates the probability that the display D_t will cause a purchase in the future S. The display valuation then becomes

$$CPA \cdot \alpha(x_t) \cdot \mathbb{E}(S | X_t = x_t, D_t = 1).$$

In this work, we make the common approximation that having no click on display D_t has no more effect on the user purchase probability than showing no display at all. As a consequence, we estimate this factor $\alpha(x_t)$ as a function of the expected number of purchases given that the displays D_t is clicked $C_t = 1$ or not $C_t = 0$:

$$\alpha(x_t) = 1 - \frac{\mathbb{E}(S | C_t = 0, X_t = x_t, D_t = 1)}{\mathbb{E}(S | C_t = 1, X_t = x_t, D_t = 1)}.$$

It equals 1 if the sales are fully incremental, namely, the user would not have purchased without the click and 0 if there is no incrementality, that is if the click did not increase the sales probability.

In this work, we first review the literature existing on RTB and exhibit its limitations for display valuation in Section 2. Then, the technical contributions we present are the following:

(1) *Statement of the bidder problem.* In Section 3, we provide a detailed description of the bidder's problem and the hypotheses we need to turn it into an optimization procedure solved with classic statistical tools. From there, we challenge the greedy formula and state why, to maximize its revenue, a bidder should consider the expected number of *additional* sales ΔS this display might cause in the future rather than the expected number of sales S_t that will be attributed to this display.

(2) *A model using the causal effect of clicks.* In Section 4, we introduce a new model to estimate the display valuation in line with the optimization procedure stated in the previous section. This model is based on the causal effect of clicks on sales modeled with the factor $\alpha(x_t)$. It relies on clearly stated hypotheses on the causal links between displays and sales.

(3) *A metric to measure incrementality.* In Section 5, we propose a new metric meant to evaluate how well a model predicts the expected number of *additional* sales ΔS. This measure of incrementality is complex since for a display opportunity, we cannot observe both the number of sales obtained with and without showing a display.

(4) *Reproducible and online experimental results.* Finally, in Section 6 we present reproducible offline experiments on a public data set from Criteo, a large Demand Side Platform that allows its clients to externalize the bidding process. We show interesting qualitative results for the factor $\alpha(x_t)$ as well as quantitative improvements over the greedy formula measured with the incrementality metric. We also provide online results we have obtained online in a large scale production environment. While our methodology is simple, our experimental results indicate that it is robust enough to allow for the theoretical assumptions and implementation choices we made.

2 RELATED WORK AND CONTRIBUTION

Real Time Bidding (RTB) has been the standard for selling ad inventory on the Internet for almost one decade, fueling an extensive literature [Choi et al., 2020, Wang et al., 2017] on online bidding. In the early stages, the bidders were typically getting revenue from *clicks*, and much work was thus done on click prediction models [Chapelle et al., 2015, McMahan et al., 2013].

Clicks however are not the ultimate goal of advertising. What matters the most is to increase the number of *sales* on the advertiser's website. To bridge the gap between clicks and sales, one may rely on heuristics to determine which clicked display, if any, would have caused a sale. On one hand, *last click* – the most common attribution rule – states that a sale should be *attributed* to the last click preceding it. On the other hand, several more sophisticated attribution rules rely on advanced Machine Learning models [Dalessandro et al., 2012, Ji et al., 2016, Singal et al., 2019, Zhang et al., 2014]. Bidders are then incentivized on *attributed sales*, so that it is now usual to estimate the value of a display as "Cost Per Action (CPA) × Probability that the display will receive an attribution". A game theoretical analysis of the situation is provided in [Berman, 2018].

Over the decade, the bid formula shifted from this display value estimate for several reasons. First, the market globally moved from second-price auctions to first-price auctions [Despotakis et al., 2019, Heymann, 2020]. The bidder now needs to estimate the distribution of the price to beat – the highest bid of the competition – to compute its optimal bid [Krishna, 2009]. Second, it is common for the bidder to have some constraints on the advertising campaign, such as a maximum budget per day, or a maximum cost per click [Conitzer et al., 2018, Heymann, 2019]. While in several cases a linear scaling applied to the display valuation may be enough to optimally satisfy a budget constraint, the value of this scaling factor is not known in

advance, and several articles propose to update the bidding strategy to better take such constraint into account [Cai et al., 2017, Grislain et al., 2019, Lee et al., 2013, Yang et al., 2019]. In the present work, we do not consider such constraints, but those methods might be applied on top of ours. Finally, bidding with the probability that a sale is attributed to current display may not be optimal when there are several display opportunities on the same user. The authors of [Diemert et al., 2017] present a simple but efficient heuristic for taking into account user sequences with several clicks, explicitly lowering the bids right after a click. The idea that showing several displays in a row does not serve the advertiser well is at the origin of several works on pacing and probabilistic throttling. While an analytical solution of the pacing problem for display advertising is presented in [Fernandez-Tapia, 2019], most industrial solutions rely on heuristics [Agarwal et al., 2014, Chen et al., 2011, Lee et al., 2013, Xu et al., 2015].

We propose in Section 3 a formulation of the bidder problem on the sequence of bid requests for one user, and retrieve the intuitive result that we should account for the probability that the display caused the attributed sale. Our formulation is closely inspired by Reinforcement Learning [Sutton et al., 1998], defining the sequences of requests for one user as an episode. While Reinforcement learning has already been proposed to improve the bidder in [Cai et al., 2017], previous approaches only encoded the remaining budget into the state, assuming i.i.d. requests.

Closely related to the sequentiality of the interactions with the user, the question of incrementality is an attractive field of research [Lewis and Rao, 2015, Lewis et al., 2011]. Incremental sales are those that needed the display to happen and may be measured with an A/B test by switching the bidder off on a part of the population. As argued in [Xu et al., 2016], it would be in the interest of the advertiser that the bidders value a display with the *lift*, which we may understand as the probability that the display cause the sale. The difficulty here is that we never directly observe incremental sales. We either observe a sale on a user after a sequence of displays, or – in the case of an A/B test of incrementality – after no display at all. This is a typical case of a causal inference problem, where we want to estimate the effect of a treatment (here bidding to buy display ads) on an outcome (in this case the sale).

Several works have already proposed bidding methods optimized for incremental sales. For instance, [Diemert et al., 2018] proposed to estimate the causal effect on sales of bidding on a user, by applying randomized bid factor on each user. This allows learning models predicting which users are impacted by display ads. However, if those users are seeing several displays, as it is typical, it does not allow retrieving which of those displays caused the sale. Similar to our work [Rahier et al., 2020] leverage the way the causal effect is mediated to derive an uplift estimate. A method to decrease the noise of incremental sales measurement is presented in [Johnson et al., 2017]

Some authors also proposed to use observational data to estimate the causal impact of a display. Observational studies typically require to assume that there are no unobserved confounders; but because of the large state space they usually also need additional assumptions to lower the variance. [Xu et al., 2016] propose to sample user states of random times, and instead of looking if a specific display is won, they rather check if there is at least one won display

in the following time window. The (implicit) hypotheses here seems to be that the display only impact the sale through the "number of displays in a time window" variable. Similarly [Moriwaki et al., 2020] propose to unbias a model estimating the probability that a user is led to buy after a request by using inverse propensity weights, but also summarize the impact of winning a display by the increase of the variable "number of displays". Such hypothesis seems imperfect, as variables such as the size of the display or the quality of the publisher are known to have a huge impact at least on clicks or attributed sales; and it seems therefore likely that they also have great impact on incremental sales. By contrast, our methodology allows taking into account all the information available on the user and the display opportunity. Finally, another issue of bidding for incremental sales is that the bidder is not paid for incremental sales.

Our aim is to show that tools from causal reasoning [Pearl, 1995, 2009, Peters et al., 2017] can indeed improve the performance of a bidder. However, what we propose is slightly different from the above-mentioned literature: we use causal methods to increase the performance of a bidder *measured with an attribution rule* instead of being measured as the total number of sales. The motivation behind focusing on the received attribution is threefold: (a) the attribution is still the main metric in the industry, which is something that might persist because of the business models of digital advertising, (b) our approach can ingest causally motivated attribution as input, which is currently an active track of research [Dalessandro et al., 2012, Singal et al., 2019], (c) from a practical and methodological perspective, it is easier to work with attributed sales – for which we have methodologies, models and benchmarks – than with incremental sales, so while the models to address the two questions are similar, it is much easier, to test, measure, compare results and improve with attributed sales than it is with incremental sales.

This paper assumes a given attribution mechanism but does not challenge it. To better align bidders and advertisers objectives, a better attribution, based on causality, seems to be necessary.

3 THE BIDDING PROBLEM

In this section, we propose a formalization of the bidder's problem trying to maximize its revenue. In particular, we carefully consider the fact that a bidder will have a sequence of opportunities to display ads to the same user. Our framework takes as an input the compensation mechanism chosen by the advertiser and is general enough to apply to a bidder paid for either attributed sales or incremental sales. The main result of this section is Theorem 3.1, which states that a display valuation should be the difference of two terms:

- Its impact on the expectation of additional attributed sales.
- And its impact on expected cost paid by the bidder at later time steps.

The bidder should then find the bid which maximizes the expected difference of display value and cost on the current auction. For example, when the auction is second price, it must bid directly with the display value. While intuitively simple, this result allows us to cleanly separate the display valuation from all the issues related to the type of auction. The correct mathematical definition of those terms requires a few technical hypotheses. We decide to state these

hypotheses explicitly while they are often implicitly used by most authors.

3.1 Mathematical formulation

When a user browses the internet, he encounters pages with ad banners to display. Each of these will trigger an auction, where advertisers may bid to buy the display opportunity. The highest bidder will then show a display to the user, and pay a price to the auctioneer. From the point of view of a bidder, the sequence of bid requests for a given user defines a stochastic process, where at each time step t:

(1) A bidder receives a bid request containing the *state X_t* of the user. It encodes everything relevant on this user and request, such as the user past interactions with the partner website, the timestamp, the displays made at previous time steps, the current website from which the request is received, the size of the ad on this page, etc.
(2) The bidder outputs a *bid b_t* for this auction based on its own valuation of the display opportunity.
(3) It then receives a binary variable D_t telling if it did *win the display* (in which case $D_t = 1$) or not (and then $D_t = 0$), and the associated *cost $Cost_t$* (the cost is by definition 0 if $D_t = 0$).
(4) If the bidder wins, it can also later observe the click C_t, defined as 1 if the *display is clicked* and 0 if it is not. Note that clicks are not what the bidder optimizes for, so they could be ignored to define the bidder's problem. Still, we will use clicks in the model of Section 4.

During the sequence, the user may buy some items, and at the end of the sequence, the advertiser may decide to attribute some of those sales to the bidder. The bidder thus observes the number of *attributed sales S* at the end of the sequence, and receives a *payment $S \cdot CPA$* where CPA is a constant defining the value (in euro) of an attributed sale. The bidder processes in parallel the sequences of the thousand or sometimes millions of users in its campaign. To respond to all these requests, the bidder chooses a *bidding policy π*, which is formally a mapping from the space of states to distributions on bids (so the bid at time t b_t is sampled from $\pi(x_t)$). The bidder maximizes its expected payoff:

$$\mathbb{E}\left[S \cdot CPA - \sum_{t \in sequence} Cost_t\right].$$

Note that this expectation depends on π, even if we do not write it explicitly to simplify the notations.

3.2 Assumptions

In order to derive a bidding formula from this model we formulate a series of assumptions.

ASSUMPTIONS 1 (FULLY OBSERVABLE MDP). *The state X_t is fully observable, and the sequence (X_t, b_t) of states and bids on a given user forms a Markov Decision Process (MDP).*

ASSUMPTIONS 2 (INDEPENDENT USERS). *Each user is an independent instance of this MDP.*

With the first two hypotheses, the proposed framework would become an instance of a Reinforcement Learning (RL) problem,
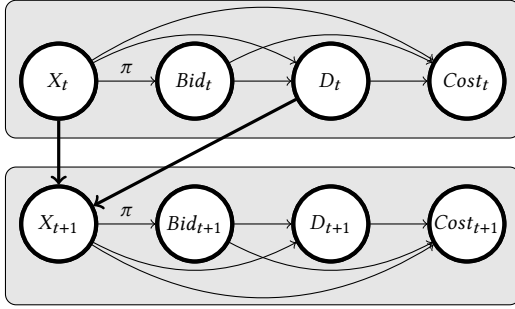
Figure 1: Causal graph of the bidding process. The user state X_t is modeled as a Markov Chain, where the transition matrix from X_t to X_{t+1} only depends on D_t, the variable that encodes the realisation of a display at time t.

where we would have one trajectory per user, the continuous actions would be the bids b_t, the reward at step t would be $-Cost_t$, and the sequence ends with a final reward $S \cdot CPA$.

ASSUMPTIONS 3 (BID IMPACTS FUTURE ONLY THROUGH THE DISPLAY). *The next state X_{t+1} only depends on X_t and D_t, not on the bid b_t.*

Note that the cost $Cost_t$ and the display D_t depend on the bid; this assumptions states that both the bidder and the environment forgets about the exact bid and cost at time t after observing D_t. Formally, this means that, for any policy π of the bidder, the sequence of variables (X_t, D_t) forms a Markov chain. The causal graph of Figure 1 summarizes those hypotheses.

Those assumptions may not be fully satisfied in practice. For example:

- Competitors might have some private information on the user. Since display winning rate and final sale may depend on this private information, that would make Assumption 1 wrong.
- Assumption 3 would be violated if the auctioneer uses the bid at time t to compute a reserve price on a next auction for the same user at time $t + k$; or if the bidder itself decides to modulate its next bids as a function of $Cost_t$.

Despite those possible shortcomings, we believe those assumptions to be reasonable. We would like also to note that *they are usually implicitly assumed when the display is valued with a function depending only on the user state x_t.* To our knowledge this includes most (if not all) prior work which values a display with an estimated click-through rate or conversion rate depending only on x_t.

3.3 Shortcomings of the greedy policy

As discussed before, the advertiser typically attributes the sales with some simple rule such as attributing to the last click of the user. Such a rule specifies exactly to which display the sale is attributed. This allows to *naturally* replace the final attributed sale S on the whole sequence of events by a sequence of attributed sales S_t defined at each time step: S_t is 1 if a sale is attributed to the display at time t, and 0 otherwise. In this setting, we define $S \triangleq \sum_t S_t$, and the goal is still to maximise $CPA \cdot S - \sum_t Cost_t$. Note however that

the variable S_t is only observed at the end of the sequence, when the advertiser provides the attribution, and not immediately after the action (bid) at time t. In particular, if the user is shown several displays, S_t *might be impacted by displays at later time steps*.

We now define a greedy bidder as a bidder that maximizes, at each time step, the expected payoff at *this* time step, $\mathbb{E}[CPA \cdot S_t - Cost_t]$. For example, on a second price auction, it is well-known that the solution is then to bid its expected gain[1]:

$$b_t = CPA \cdot \mathbb{E}(S_t | X_t = x_t, D_t = 1)$$

While not optimal for the whole sequence, the greedy bidder is still widely used in practice due to its simplicity.

We can now explain why the greedy policy is not optimal. An attribution such as last click is not perfect. For instance, suppose that a first display is sufficient to cause a sale that is yet to come. If we show a second display before, it will receives the attribution that the first display would have received if we skipped the second display. However, in total the bidder will get the same reward. The second display is useless but it receives all the credit and will be learned as more valuable than the first one.

3.4 Optimal policy

Next we derive the optimal bid at a time τ, assuming we know that the bidder will follow a predefined policy π in the future. We define the future cost $FCost_\tau \triangleq \sum_{t > \tau} Cost_t$, and the impact of the display on the expected reward

$$\Delta S(x_\tau) \triangleq \mathbb{E}[S | X_\tau = x_\tau, D_\tau = 1] - \mathbb{E}[S | X_\tau = x_\tau, D_\tau = 0] \quad (1)$$

and on the future costs

$$\Delta FCost(x_\tau) \triangleq \mathbb{E}[FCost_\tau | X_\tau = x_\tau, D_\tau = 1] - \mathbb{E}[FCost_\tau | X_\tau = x_\tau, D_\tau = 0]. \quad (2)$$

We can now state the main result:

THEOREM 3.1 (OPTIMAL BID AT TIME STEP τ). *Under Assumptions 1, 2 and 3, the optimal bid at time τ is*

$$b_\tau^* \triangleq argmax_b \mathbb{E}[D_\tau \cdot (V(x_\tau) - Cost_\tau) | Bid_t = b]$$

with the display valuation defined as

$$V(x_\tau) \triangleq CPA \cdot \Delta S(x_\tau) - \Delta FCost(x_\tau).$$

We defer the proof of this Theorem to Appendix A. Note that the expectation in the formula of Theorem 3.1 is only on the randomness of the auction at time t. The bidder should set its bid as a function of the valuation $V(x_\tau)$ following auction theory [Krishna, 2009]. This bid adjustment in the specific context of RTB has been studied in [Cai et al., 2017, Despotakis et al., 2019, Grislain et al., 2019, Lee et al., 2013, Yang et al., 2019] and is not in the scope on this work. For example, in the simple setting of second price auction, the optimal bid is deterministic, $b^* = V(x_\tau)$. We will now focus on estimating the valuation $V(x_\tau)$.

---

[1]Note that this expectation actually depends on the policy of the bidder after time t, because S_t may be impacted by future displays. In practice, this technicality is ignored, and the expectation is estimated with data generated with a previous version of the bidder policy.

4 MODELING THE CAUSAL EFFECT OF CLICKS

According to Theorem 3.1, we need to estimate the impact of the display on the additional attributed sales $\Delta S(X_\tau)$ and future cost $\Delta FCost(X_\tau)$. In this paper, we propose a new method to estimate $\Delta S(X_t)$ by leveraging the randomness of clicks. As we believe the biggest source of inefficiency in the greedy bidder is the attribution of sales to the last click, in this work we will only focus on estimating the term $\Delta S(X_t)$ and approximate the term $\Delta FCost(X_\tau)$ by 0 in our experiments.

Let us notice that for a given sample, we observe either what happens after $D_t = 1$ or after $D_t = 0$, but never both: we need here to estimate the causal effect of the display D_t on the reward in Equation (1). Although we assume that all variables are observed, this may still be challenging in practise due to the large dimension of the state space, and the high probability that $S = 0$.

In this section, we sometimes drop the index t for simplicity, as we now only care about variables at time of the bid. Moreover, we assimilate binary random variables such as D and C to the events $D = 1$, $C = 1$. Thus we write $P(C|D, X=x)$ for of $P(C_t=1|D_t=1, X_t=x)$; and we note $\overline{D}, \overline{C}$ the events $D = 0$ or $C = 0$. Finally, since in practice having several sales is very rare, we assume that S is binary, and thus note S the event $S > 0$. This also allows us to fit S with a logistic regression. Hence Equation (1) becomes

$$\Delta S(x) = \mathbb{P}(S|X = x, D) - \mathbb{P}(S|X = x, \overline{D}).$$

This quantity can be understood as "the causal effect of display D on the attributed sale S". We are thus trying to retrieve which display(s) caused the sale attribution to the bidder. Also, because a user is typically exposed to dozens of displays on a single day and a sale is a rare event, the dilution of the sale signal prevents the use of directly fitting the model $\mathbb{P}(S|X = x, D)$. We thus rely on a few more assumptions.

4.1 Assumptions

A common assumption used implicitly by most attribution systems, is that non clicked displays have no impact on sales. This greatly reduces the number of displays one should consider when trying to find "which display might have caused the sale".

ASSUMPTIONS 4 (No POST DISPLAY EFFECT.). *The system is well represented by the causal graph in Figure 2. In particular, C_t blocks all directed paths starting from D_t and finishing to S.*

In other words, displays may only cause a sale through a click. We now have:

$$\mathbb{P}(S|X = x, \overline{D}) = \mathbb{P}(S|X = x, \overline{D}, \overline{C}) = \mathbb{P}(S|X = x, D, \overline{C}).$$

The first equality comes from the fact that a click cannot occur without a display. The second equality can be derived using Pearl's d-separation, after observing that $S \perp\!\!\!\perp D|(X, C)$, where $\perp\!\!\!\perp$ is the d-separation symbol.

REMARK 1 (DISCUSSION ON ASSUMPTION 4).

- *The no post display effect assumption is quite commonly assumed in the industry when attributing the sales. For example, in the case of a "last click" attribution, only clicked displays can be credited for a sale which implicitly implies the not post*
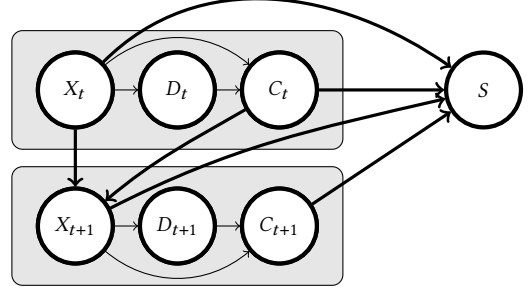

Figure 2: Causal Graph of the effect of the user state X_t on its purchase S through the bidding process. Comparing with Figure 1, bid and cost are omitted for clarity and we add the Assumption that the effects of the display D_t on the sale S and the next user state X_{t+1} are mediated by the click C_t. In other words, we assume that only a clicked display may have an impact on the probability of seeing a sale.

---

display effect assumption. While it might still be possible that a non-clicked display would cause the sale and that then another clicked display would make the sale to be attributed to the bidder, we believe it is reasonable to assume that attributed sales are caused only by clicked displays.
- *Note that in this model, the effect on displays D_t to the next state X_{t+1} is fully mediated by the click C_t. For example, this implies that the number of past displays cannot be part of the state. Since this is a feature often used by bidders, we recognize that it indicates a potential statistical relevance not taken into account in our framework.*

4.2 The incremental bidder

Since the click C is binary, we can always decompose the probability of having a sale after a display $\mathbb{P}(S|D, X=x)$ into

$\mathbb{P}(S|D, X=x)$

$= \mathbb{P}(S|C, D, X=x) \cdot \mathbb{P}(C|D, X=x) + \mathbb{P}(S|\overline{C}, D, X=x) \cdot \mathbb{P}(\overline{C}|D, X=x)$

$= \mathbb{P}(C|D, X=x) \cdot \Big(\mathbb{P}(S|C, D, X=x) - \mathbb{P}(S|\overline{C}, D, X=x)\Big) + \mathbb{P}(S|\overline{C}, D, X=x).$

This leads to

$$\Delta S(x) = \mathbb{P}(C|D, X=x) \cdot \Big(\mathbb{P}(S|C, D, X=x) - \mathbb{P}(S|\overline{C}, D, X=x)\Big)$$
$$+ \Big(\mathbb{P}(S|\overline{C}, D, X=x) - \mathbb{P}(S|\overline{D}, X=x)\Big).$$

Following our previous assumptions, the last term $\mathbb{P}(S|\overline{C}, D, X=x) - \mathbb{P}(S|\overline{D}, X=x)$, which we may interpret as the "pure post display effect of D on the sale", is assumed to be 0. Hence

$$\Delta S(x) = \mathbb{P}(C|D, X=x) \cdot \Big(\mathbb{P}(S|C, D, X=x) - \mathbb{P}(S|\overline{C}, D, X=x)\Big). \quad (3)$$

All the terms in this formula may be estimated directly from the data, using standard supervised learning methods such as logistic

regression. With a simple rewriting, if $\mathbb{P}(S|C,D,X{=}x) > 0$, Equation (3) becomes

$$\Delta S(x) = \mathbb{P}(C|D,X{=}x) \cdot \mathbb{P}(S|C,D,X{=}x) \cdot \left(1 - \frac{\mathbb{P}(S|\overline{C},D,X{=}x)}{\mathbb{P}(S|C,D,X{=}x)}\right), \quad (4)$$

where we call the last term the *incrementality factor*. It is usually between 0 and 1 and we might think about it as "the probability that the click caused the sale". It equals 1 if the sale is fully incremental, namely, it could not have occurred without the click: $\mathbb{P}(S|\overline{C},D,X) = 0$. On the contrary it is valued 0 if there is no incrementality, namely if the click did not increase the sales probability: $\mathbb{P}(S|\overline{C},D,X) = \mathbb{P}(S|C,D,X)$. While it could in theory be negative (if the click caused the user not to buy) it remains rare and in such a case we may simply submit a bid of 0, which is enough to ensure we loose the display. We clarify that the bidder is *incremental* in the sense that it bids the lift of attributed sales – or whatever is measured by the variable S: if S was actually standing for the total number of sales, the bidder would optimize for the incremental sales.

REMARK 2 (DIFFERENCE BETWEEN GREEDY AND INCREMENTAL BIDDERS). *With a greedy bidder, a display is typically valued $CPA \cdot \mathbb{P}(C_t|D_t,X{=}x) \cdot \mathbb{P}(S_t|C_t,D_t,X{=}x)$. The proposed model makes two changes compared to this greedy bidder:*

(1) *We multiply by the incrementality factor from Equation (4).*
(2) *We replace S_t, the sale attributed to display D_t by S (sales attributed to the whole sequence), in the $\mathbb{P}(S|C,D,X{=}x)$ part of the model. We pinpoint that our model does not depend on the method used to attribute sales to displays (typically last click) as long as it is attributed to the displays sequence made to the user.*

5 AN INCREMENTALITY METRIC

We are interested in valuing more the displays that lead a user to make a purchase in the future while he would not have not without. Namely, we would like to evaluate how good our model is at predicting $\Delta S(x)$. Given Equation (3), $\Delta S(x)$ can be split into two parts, first $\mathbb{P}(C|D,X{=}x)$ that is a quite mature model in the industry and a second model $\mathbb{P}(S|C,D,X{=}x) - \mathbb{P}(S|\overline{C},D,X{=}x)$ that we have introduced in Section 4.2. Since we cannot observe both the clicked display and its counterfactual, assessing the quality of such a model offline is a complex task. In this Section, we build a random variable Y such that

$$\mathbb{P}(Y|C,D,X{=}x) = \mathbb{P}(S|C,D,X{=}x) - \mathbb{P}(S|\overline{C},D,X{=}x)$$

that we cannot observe. However, we introduce a metric that is able to evaluate the capacity of our model to predict this variable Y while observing only the click C and sales S outcomes.

5.1 The generative model

As in Section 4.2 we consider only won displays, that are described by a context X, and have two observable effects: they can be clicked C or not $\overline{C}$ and they can be followed by a sales S or not $\overline{S}$. Intuitively, we would like to divide those displays in four classes, which we call their *display type* $T \in \{a,n,y,d\}$:

(1) The displays that **always** lead to a sale whether they are clicked or not : $T = a$.

Table 1: Our generative model assumes that, given the display type ($T = a$, n or y) and whether it is clicked C or not $\overline{C}$, we can determine if it leads to a sale S or not $\overline{S}$.

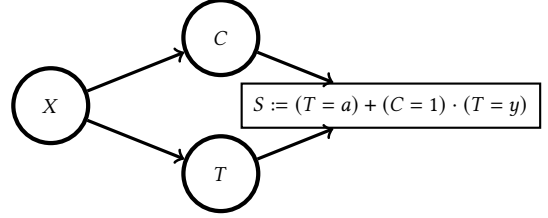| $\cap$	$T = a$	$T = n$	$T = y$
$\overline{C}$	S	$\overline{S}$	$\overline{S}$
C	S	$\overline{S}$	S

Figure 3: Causal graph of the generative model. The display context X influences the display type $T \in \{a,n,y,d\}$ (specifically if it is incremental Y) and whether the display is clicked C. Then, T and C determine if this display will leads to a sale S following Table 1. Note that there is no causation from C to Y to illustrate the independence assumption of Proposition 5.1.

(2) The displays that **never** lead to a sale whether they are clicked or not : $T = n$.
(3) The **incremental** displays that lead to a sale if they are clicked but do not if they are not clicked: $T = y$.
(4) The **decremental** displays that do not lead to a sale if they are clicked but do if they are not clicked: $T = d$.

ASSUMPTIONS 5 (NO DECREMENTAL DISPLAYS). *We assume for all x, $\mathbb{P}(S|C,X = x) \geq \mathbb{P}(S|\overline{C},X = x)$.*

This mean that there are no decremental displays (or equivalently, the incremental factor is positive). This assumption is quite natural in practice since we do not expect a clicked ad to discourage a user to buy a product. Of course, this display type is never observed directly, but for any given context x, we can always assume that the variable S is a function of C and an independent unobserved variable $T \in \{a,n,y\}$ as defined in Table 1 [2]. The corresponding causal graph is depicted in Figure 3.

Since we are looking for incrementality, $T = y$ is our variable of interest, even if it cannot be observed from the data. Indeed, the displays that have been clicked and then have led to a sale ($C \cap S$) and the display that were not clicked and have not led to a sale ($\overline{C} \cap \overline{S}$) might either be incremental or not. To determine if they are incremental one should have observed the counterfactual. Finally, similarly to $\overline{C}$ and $\overline{D}$ we denote by Y the random variable associated to the event $T = y$ and $\overline{Y}$ to $T \neq y$.

---

[2]For example, we could define $\mathbb{P}(T = a|X = x) \triangleq \mathbb{P}(S|X = x, \overline{C})$ and $\mathbb{P}(T = n|X = x) \triangleq \mathbb{P}(\overline{S}|X = x, C)$.

5.2 Reverted incremental likelihood

Y is a Bernoulli variable, hence if we have an estimator f that predicts Y given X its log-likelihood is

$$LLH_B(Y, f(X)) = Y \log f(X) + \overline{Y} \log(1 - f(X)).$$

Since we cannot observe Y, we cannot compute this likelihood directly from the data. However, Proposition 5.1 gives us an unbiased estimate of this likelihood based only on the observed values C and S. This unbiased estimate *reverts the label* as in [Jaskowski and Jaroszewicz, 2012] while taking into account the partial randomness of the clicks.

PROPOSITION 5.1. *With the generative model of Figure 3 and Assumption 5, the reverted incremental likelihood*

$$RLLH_B(C, S, f(X)) = \frac{C}{\mathbb{P}(C|X)} LLH_B(S, f(X)) + \frac{\overline{C} \cap S}{\mathbb{P}(\overline{C}|X)} \log \frac{1 - f(X)}{f(X)}$$

is an unbiased estimator of the direct incremental likelihood $LLH_B(Y, f(X))$, for any predictor f.

We defer the proof of Proposition 5.1 to Appendix C. Provided we have a correct $\mathbb{P}(C|X)$ model, we can thus estimate the incremental likelihood from observable data. This is interesting in practice since it allows to us to evaluate offline the incremental performance of our models to perform model selection. In the case of the incremental bidder, it is useful to assess the quality of the incremental factor as a whole instead of evaluating $\mathbb{P}(S|\overline{C}, D, X=x)$ on one hand and $\mathbb{P}(S|C, D, X=x)$ on the other while having no guarantee on the quality of their ratio. Note that this still requires a click prediction model to estimate $\mathbb{P}(C|X)$. Relying on a model to assess the performances of another is not ideal but since the click prediction models are mature in the industry and trained on much more data, we can safely use them to evaluate sales prediction models. In Appendix D, we illustrate Proposition 5.1 on a data set that we simulate such that we can observe the hidden variable T. That allows us to compute the direct incremental likelihood and to compare it the reverted one to assess the performance of the incremental bidder.

6 EXPERIMENTAL RESULTS

6.1 Offline analysis

In order to evaluate our methodology, we run experiments on the Criteo Attribution Modeling for Bidding public data set [Diemert et al., 2017]. This data set consists in 16 million displays sent by Criteo to 6 million users over a period of 30 days. To each display is associated a set of context features (that we have denoted by X_t so far), if the display was clicked (denoted by C_t so far), and if it has led to a sale that was attributed to Criteo (denoted by S_t so far). For simplicity, we assume the value of an attributed sale (denoted by CPA) to equal 1. As a baseline, we train the greedy bidder that values a display with context features x by $\mathbb{P}(C_t|X=x) \cdot \mathbb{P}(S_t|C_t, X_t=x)$ (see Remark 2). Note that the context features are hashed [Weinberger et al., 2009], and each context x ends up being represented by sparse vector of dimension 2^{16}. Then, since we have access to the user identifier, we can reconstruct the user timelines. Specifically, we can determine if the sequence of displays has led to an attributed sales (denoted by S so far). With this label, we train the incremental
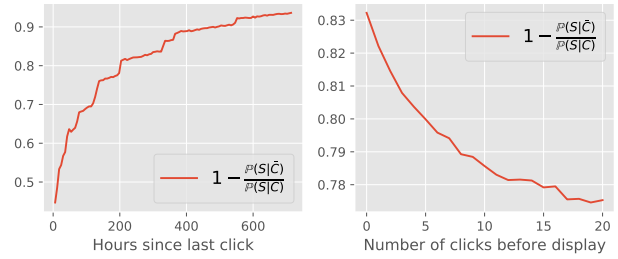

Figure 4: Value of the incremental factor given the number of hours elapsed since the latest click. As our intuition would suggest, the most recent the click is, the less incremental the display is expected to be. Similarly, if a user has already clicked on several displays then the following clicked displays are less likely to be incremental.

bidder derived in Section 4.2 and compute the incremental factor from Equation (4) [3].

In Figure 4, we report the average value of this factor over the displays that we have grouped in two manners, first given the time elapsed since the latest user's click, and second, given the number of clicks already made by the user in the timeline. In both cases, we observe that the incremental factor illustrates what we could have foreseen: the more recent the click is, the less incremental the display is expected to be and, similarly, if a user has already clicked on several displays then the following clicked displays are less likely to be incremental. Note that the first relationship –between the incrementality and the time elapsed since the latest click– is at the heart of the work in [Diemert et al. 2017, Figure 5] where they model it with one minus a decreasing exponential parameterized with a decay factor λ. It is interesting to see that the shape is similar to what we observe with the incremental factor but also that our approach learns a clean monotonic relationship with a non parametric model (namely we do not specify that we look for a decaying effect parameterized with a factor λ to learn). Also, our incremental bidder is not limited to the time elapsed since the latest click but can learn from all the available features (such as number of clicks before display for example).

On the same data set, we compare the performances of the greedy and the incremental bidders evaluated with the incremental likelihood computed thanks to Proposition 5.1. In both cases, we consider that the click prediction is evaluated separately and we isolate the part of the formula that gets multiplied by $CPA \cdot \mathbb{P}(C_t|D_t, X=x)$. That is, for the greedy bidder we use $\mathbb{P}(S_t|C_t, D_t, X=x)$ as predictor of Y and $\mathbb{P}(S|C_t, D_t, X_t=x) - \mathbb{P}(S|\overline{C_t}, D_t, X_t=x)$ for the incremental bidder. We train the models on 80% of the users and keep the remaining 20% for a test set. We report the results in Figure 5 where we varied the size of the hashing space from 2^{10} to 2^{16}. We observe that both models obtain better performances when their predictive capacities (induced by the feature space's size) are increased. Also, we observe that the incremental bidder obtains better results than

---

[3] Note that all the experiments conducted on this data set are hosted on Google colab https://colab.research.google.com/github/incrementality-factor-submission/ incrementality-factor/blob/main/incrementality_factor.ipynb
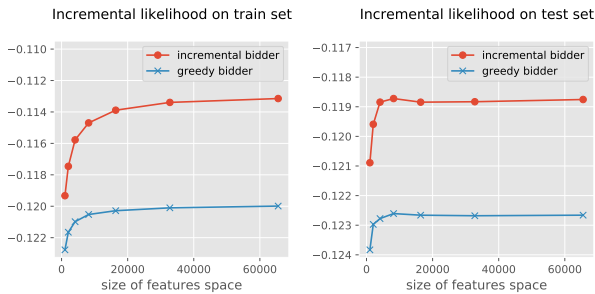
Figure 5: Value of the incremental likelihood for our bidding formula compared to the baseline, i.e. last click model. We plot it in regards to the size of the features space that quantifies the capacity of the model.

the greedy bidder. It reaches better performances than the greedy bidder with 1,000 features instead of 60,000.

We have endorsed the incrementality log-likelihood with an experiment in a simulated environment for which we were able to recover the generative parameters. Then, we have shown that the incremental bidder manages to improve significantly this metric over the greedy bidder on a public data set. However, the ground truth effect of such changes can only be known through online experiment, by nature not reproducible. It is the not reproducible nature of the problem that justifies the need for offline metrics.

6.2 Live experiment

Finally, we explain how we modify a production bidding algorithm to test our model online. The baseline is made of (1) a module that predicts the probability that an opportunity, if won, will be attributed a sale with the last click rule (the greedy bidder), and (2) a module that shades this prediction to account for the first-price auction rule. Our change simply consists in (a) learning $\mathbb{P}\big(S|\overline{C}, D, X\big)$ and $\mathbb{P}\big(S|C, D, X\big)$ with a logistic regression, (b) form the factor

$$\alpha(X) = 1 - \frac{\mathbb{P}\big(S|\overline{C}, D, X\big)}{\mathbb{P}\big(S|C, D, X\big)},$$

(c) multiply the production module of (1) by this *incremental factor*. Many feature engineering decision where taken in step (2b), among other, we decided to learn $\alpha(X)$ with a logistic regression.

We tested the change on and obtained a 7.8% decrease of spend for the same campaign performance in attribution on a volume worth 7,000,000 euros on the test period, which is something remarkable in the industry.

7 DISCUSSION

We introduce a reinforcement learning formulation of the bidding problem in display advertising. We derived a strategy thanks to a causal reasoning approach. The main assumption is to suppose that only clicked displays bring additional reward. We also provide an offline metrics to sanitize the approach. While the framework could be used for incremental sales, we tested it on attributed sales for the reasons given in the introduction (it is the main metric in

the industry and it is methodologically easier to test and develop on attribution).

On the experimental and implementation aspects, further work could be done, in particular to take into account the evolution of the expected future cost. We believe our framework is quite general and could be adapted to other contexts with clicks, displays and actions, in particular marketplace and recommender systems.

From a methodological perspective, we think that attribution can be useful to understand, test and improve causal models (which can later be adapted for incremental sales). From an industry wise perspective, our approach could be combined with incremental attribution framework such as [Singal et al., 2019]. Such a decomposition offers several advantage as the role of the different agents are clearly defined.

REFERENCES

Deepak Agarwal, Souvik Ghosh, Kai Wei, and Siyu You. 2014. Budget Pacing for Targeted Online Advertisements at LinkedIn. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*. Association for Computing Machinery, New York, NY, USA, 1613–1619. https://doi.org/10.1145/2623330.2623366

Ron Berman. 2018. Beyond the last touch: Attribution in online advertising. *Marketing Science* 37, 5 (2018), 771–792.

Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. 2017. Real-Time Bidding by Reinforcement Learning in Display Advertising. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)*. Association for Computing Machinery, New York, NY, USA, 661–670. https://doi.org/10.1145/3018661.3018702

Olivier Chapelle, Eren Manavoglu, and Romer Rosales. 2015. Simple and Scalable Response Prediction for Display Advertising. *ACM Trans. Intell. Syst. Technol.* 5, 4, Article 61 (Dec. 2015), 34 pages. https://doi.org/10.1145/2532128

Ye Chen, Pavel Berkhin, Bo Anderson, and Nikhil R. Devanur. 2011. Real-Time Bidding Algorithms for Performance-Based Display Ad Allocation. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. Association for Computing Machinery, New York, NY, USA, 1307–1315. https://doi.org/10.1145/2020408.2020604

Hana Choi, Carl F. Mela, Santiago R. Balseiro, and Adam Leary. 2020. Online display advertising markets: A literature review and future directions. *Information Systems Research* 31, 2 (2020), 556–575.

Vincent Conitzer, Christian Kroer, Eric Sodomka, and Nicolás E. Stier-Moses. 2018. Multiplicative Pacing Equilibria in Auction Markets. In *Web and Internet Economics - 14th International Conference, WINE 2018, Oxford, UK, December 15-17, 2018, Proceedings (Lecture Notes in Computer Science)*, George Christodoulou and Tobias Harks (Eds.), Vol. 11316. Springer, 443. https://link.springer.com/content/pdf/bbm%3A978-3-030-04612-5%2F1.pdf

Brian Dalessandro, Claudia Perlich, Ori Stitelman, and Foster Provost. 2012. Causally Motivated Attribution for Online Advertising. In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy (ADKDD '12)*. Association for Computing Machinery, New York, NY, USA, Article 7, 9 pages. https://doi.org/10.1145/2351356.2351363

Stylianos Despotakis, R Ravi, and Amin Sayedi. 2019. First-price auctions in online display advertising. *Available at SSRN 3485410* (2019).

Eustache Diemert, Amélie Héliou, and Christophe Renaudin. 2018. Off-policy learning for causal advertising.

Eustache Diemert, Julien Meynet, Pierre Galland, and Damien Lefortier. 2017. Attribution Modeling Increases Efficiency of Bidding in Display Advertising. In *Proceedings of the ADKDD'17 (ADKDD'17)*. Association for Computing Machinery, New York, NY, USA, Article 2, 6 pages. https://doi.org/10.1145/3124749.3124752

Joaquin Fernandez-Tapia. 2019. An analytical solution to the budget-pacing problem in programmatic advertising. *Journal of Information and Optimization Sciences* 40, 3 (2019), 603–614.

Lauren Fisher. 2019. US Programmatic Digital Display Ad Spending. https://www.emarketer.com/content/us-programmatic-digital-display-ad-spending

Nicolas Grislain, Nicolas Perrin, and Antoine Thabault. 2019. Recurrent Neural Networks for Stochastic Control in Real-Time Bidding. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 2801–2809. https://doi.org/10.1145/3292500.3330749

Benjamin Heymann. 2019. Cost per Action Constrained Auctions. In *Proceedings of the 14th Workshop on the Economics of Networks, Systems and Computation (NetEcon '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/3338506.3340269

Benjamin Heymann. 2020. How to bid in unified second-price auctions when requests are duplicated. *Operations Research Letters* 48, 4 (2020), 446–451. https://doi.org/10.1016/j.orl.2020.05.010

Maciej Jaskowski and Szymon Jaroszewicz. 2012. Uplift modeling for clinical trial data. In *ICML Workshop on Clinical Data Analysis*, Vol. 46.

Wendi Ji, Xiaoling Wang, and Dell Zhang. 2016. A probabilistic multi-touch attribution model for online advertising. *International Conference on Information and Knowledge Management, Proceedings* 24-28-Octo (2016), 1373–1382. https://doi.org/10.1145/2983323.2983787

Garrett A. Johnson, Randall A. Lewis, and Elmar I. Nubbemeyer. 2017. Ghost Ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research* 54, 6 (Dec. 2017), 867–884. https://doi.org/10.1509/jmr.15.0297

Vijay Krishna. 2009. *Auction theory*. Academic press.

Kuang-Chih Lee, Ali Jalali, and Ali Dasdan. 2013. Real Time Bid Optimization with Smooth Budget Delivery in Online Advertising. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising - ADKDD '13*. ACM Press, New York, New York, USA, 1–9. https://doi.org/10.1145/2501040.2501979

Randall A. Lewis and Justin M. Rao. 2015. The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics* 130, 4 (Nov. 2015), 1941–1973. https://doi.org/10.1093/qje/qjv023

Randall A. Lewis, Justin M. Rao, and David H. Reiley. 2011. Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*. Association for Computing Machinery, New York, NY, USA, 157–166. https://doi.org/10.1145/1963405.1963431

H. Brendan McMahan, Gary Holt, D. Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. 2013. Ad Click Prediction: A View from the Trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. Association for Computing Machinery, New York, NY, USA, 1222–1230. https://doi.org/10.1145/2487575.2488200

Daisuke Moriwaki, Yuta Hayakawa, Isshu Munemasa, Yuta Saito, and Akira Matsui. 2020. Unbiased Lift-based Bidding System. *arXiv preprint arXiv:2007.04002* (2020).

Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika* 82, 4 (1995), 669–688.

Judea Pearl. 2009. *Causality*. Cambridge University Press. 1689–1699 pages. https://doi.org/10.1088/1751-8113/44/8/085201

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2017. *Elements of causal inference*. The MIT Press.

Thibaud Rahier, Amélie Héliou, Matthieu Martin, Christophe Renaudin, and Eustache Diemert. 2020. Individual Treatment Effect Estimation in a Low Compliance Setting. arXiv:stat.ML/2008.03235

Raghav Singal, Omar Besbes, Antoine Desir, Vineet Goyal, and Garud Iyengar. 2019. Shapley Meets Uniform: An Axiomatic Framework for Attribution in Online Advertising. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1713–1723. https://doi.org/10.1145/3308558.3313731

Richard S. Sutton, Andrew G. Barto, et al. 1998. *Introduction to reinforcement learning*. Vol. 135. MIT Press Cambridge.

Jun Wang, Weinan Zhang, and Shuai Yuan. 2017. *Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting*. Now Publishers Inc., Hanover, MA, USA.

Kilian Weinberger, Anirban Dasgupta, John Langford, Alex Smola, and Josh Attenberg. 2009. Feature Hashing for Large Scale Multitask Learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*. Association for Computing Machinery, New York, NY, USA, 1113–1120. https://doi.org/10.1145/1553374.1553516

Jian Xu, Kuang-chih Lee, Wentong Li, Hang Qi, and Quan Lu. 2015. Smart Pacing for Effective Online Ad Campaign Optimization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, NY, USA, 2217–2226. https://doi.org/10.1145/2783258.2788615

Jian Xu, Xuhui Shao, Jianjie Ma, Kuang-chih Lee, Hang Qi, and Quan Lu. 2016. Lift-Based Bidding in Ad Selection. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI'16)*. AAAI Press, 651–657.

Xun Yang, Yasong Li, Hao Wang, Di Wu, Qing Tan, Jian Xu, and Kun Gai. 2019. Bid Optimization by Multivariable Control in Display Advertising. (2019), 1966–1974. https://doi.org/10.1145/3292500.3330681

Ya Zhang, Yi Wei, and Jianbiao Ren. 2014. Multi-touch Attribution in Online Advertising with Survival Theory. *Proceedings - IEEE International Conference on Data Mining, ICDM* 2015-Janua, January (2014), 687–696. https://doi.org/10.1109/ICDM.2014.130

Appendices

A PROOF OF THEOREM 3.1

First we pinpoint that Equations (1) and (2) only refer to observational data. The following result, that relies on using Pearl's *do operator* [Pearl, 1995, 2009, Peters et al., 2017], tells us that we can nonetheless interpret them as interventional quantities.

PROPOSITION A.1.
$$\Delta S(x_\tau) = \mathbb{E}[S|X_\tau{=}x_\tau, do(D_\tau{=}1)] - \mathbb{E}[S|X_\tau{=}x_\tau, do(D_\tau{=}0)]$$
and
$$\Delta FCost(x_\tau) = \mathbb{E}[FCost_\tau|X_\tau{=}x_\tau, do(D_\tau{=}1)]$$
$$- \mathbb{E}[FCost_\tau|X_\tau{=}x_\tau, do(D_\tau{=}0)]$$

PROOF. The variables S and D_τ are d-separated in a graph were the outcoming edges from D_τ are removed. We can thus apply Pearl's Action/observation exchange [Pearl, 1995] . A similar argument applies for $Fcost_\tau$ and D_τ. □

We can now prove Theorem 3.1.

PROOF. Noting that the cost paid at previous steps, $\sum_{t<\tau} Cost_t$, can now be viewed as a constant and thus removed from the optimisation problem, the bidder wants to maximize the expected payoff after a bid b_τ:
$$\mathbb{E}[CPA \cdot S - FCost_\tau - Cost_\tau|X_\tau = x_\tau, b_\tau]$$
Since the bid only impacts future through the display, when the bidder looses this auction ($D_\tau = 0$), its payoff is:
$$CPA \cdot \mathbb{E}[S|X_\tau = x_\tau, D_\tau = 0] - \mathbb{E}[FCost_\tau|X_\tau = x_\tau, D_\tau = 0]. \quad (5)$$
While if he wins ($D_\tau = 1$), he would receive
$$CPA \cdot \mathbb{E}[S|X_\tau = x_\tau, D_\tau = 1] - \mathbb{E}[FCost_\tau|X_\tau = x_\tau, D_\tau = 1] \quad (6)$$
and pay the cost $Cost_\tau$. Assumption 3 implies that expectations on the future of the sequence after a display (or after no display) of the quantities in Equations (5) and (6) only depend on the policy of the bidder after time t, not on its exact bid at time t. Up to a constant term, the payoff of the bidder is thus $D_\tau \cdot (V(x_\tau) - Cost_\tau)$. □

B POLICY ITERATION AND CONVERGENCE

While we are not strictly in a Reinforcement Learning setup, we may define a policy iteration, which converges to the optimal policy [Sutton et al., 1998] under mild technical assumptions[4].

To be more specific: we have only derived the optimal bid at some time step τ, when the policy π at the next time steps is frozen. From this, we then build a new policy: at every time step, answer with the bid which would be optimal if π was used in the future. A recursion argument shows that this policy is an improvement (i.e. has a higher expected return) on π at every state. Theoretically, we can then define a sequence of policies by iterating this policy improvement process, converging to an optimal policy –which can be shown with a recursion on the length of the sequence when those sequences are of bounded length– exactly as in the usual case of policy iteration in Reinforcement Learning.

---
[4]It is sufficient to assume a finite upper bound on the length of sequences.

C PROOF OF PROPOSITION 5.1

With Assumption 5, $\overline{Y}$ is the disjoint union of $T = a$ and $T = n$, hence
$$\mathbb{P}[\overline{Y}|X] = \mathbb{P}[T = a|X] + \mathbb{P}[T = n|X].$$
As the display types T are built independently of C given X, the previous equality can be written as
$$\mathbb{P}[\overline{Y}|X] = \frac{\mathbb{P}[\overline{C} \cap T = a|X]}{\mathbb{P}(\overline{C}|X)} + \frac{\mathbb{P}[C \cap T = n|X]}{\mathbb{P}(C|X)}.$$
Finally, since we can identify event $\overline{C} \cap (T = a)$ to $\overline{C} \cap S$ and $C \cap (T = n)$ to $C \cap \overline{S}$ (see Table 1), it writes
$$\mathbb{P}[\overline{Y}|X] = \frac{\mathbb{P}[\overline{C} \cap S|X]}{\mathbb{P}(\overline{C}|X)} + \frac{\mathbb{P}[C \cap \overline{S}|X]}{\mathbb{P}(C|X)},$$
and
$$\mathbb{P}[Y|X] = \frac{\mathbb{P}[C \cap S|X]}{\mathbb{P}(C|X)} - \frac{\mathbb{P}[\overline{C} \cap S|X]}{\mathbb{P}(\overline{C}|X)}. \quad (7)$$
Using this identity we compute the expectation of the incremental likelihood
$$\mathbb{E}_{X,Y}[\text{LLH}_B(Y, f(X))]$$
$$= \mathbb{E}_X\big[\mathbb{E}_Y[Y \log f(X) + \overline{Y} \log(1 - f(X))|X]\big]$$
$$= \mathbb{E}_X\big[\mathbb{P}(Y|X) \log f(X) + \mathbb{P}(\overline{Y}|X) \log(1 - f(X))\big]$$
$$= \mathbb{E}_X\Bigg[\left(\frac{\mathbb{P}[C \cap S|X]}{\mathbb{P}(C|X)} - \frac{\mathbb{P}[\overline{C} \cap S|X]}{\mathbb{P}(\overline{C}|X)}\right) \log f(X)$$
$$+ \left(\frac{\mathbb{P}(\overline{C} \cap S|X)}{\mathbb{P}(\overline{C}|X)} + \frac{\mathbb{P}[C \cap \overline{S}|X]}{\mathbb{P}(C|X)}\right) \log(1 - f(X))\Bigg]$$
Since all is conditioned on X and C and S are Bernouilli variables,
$$\mathbb{E}_{X,Y}[\text{LLH}_B(Y, f(X))]$$
$$= \mathbb{E}_{X,C,S}\Bigg[\left(\frac{C \cap S}{\mathbb{P}(C|X)} - \frac{\overline{C} \cap S}{\mathbb{P}(\overline{C}|X)}\right) \log f(X)$$
$$+ \left(\frac{\overline{C} \cap S}{\mathbb{P}(\overline{C}|X)} + \frac{C \cap \overline{S}}{\mathbb{P}(C|X)}\right) \log(1 - f(X))\Bigg]$$
$$= \mathbb{E}_{X,S,C}\Bigg[\frac{C}{\mathbb{P}(C|X)}\text{LLH}_B(S, f(X))$$
$$+ \frac{\overline{C} \cap S}{\mathbb{P}(\overline{C}|X)} \log \frac{1 - f(X)}{f(X)}\Bigg].$$

D COMPARISON OF DIRECT AND REVERTED INCREMENTAL LIKELIHOOD ON A SIMULATED DATA SET

In order to manifest the consistency of the incremental metric introduced in Proposition 5.1 and link it to the incremental factor, we generate a toy data set where the displays are described by a 2-dimensional feature vector x, allowing us to represent them on a 2d graph. Then, following our generative model, we assume that each display i, gets clicked with a probability $\psi(x_i) \in (0, 1)$ and is of type $t_i = a, y$ or c with a probability that also depends on x_i. Figure 6 shows $n = 60$ displays represented by their feature vector $x_i \in \mathbb{R}^2$ and their click c_i, type t_i and sales s_i characteristics for $i = 1, \ldots n$.

Table 3: Value of the *reverted* incremental likelihood obtained thanks to Proposition 5.1. At the expense of a wider 95% confidence interval, it converges as expected to the same value as the direct likelihood (see Table 2) without observing Y.

	$n = 100$	$n = 1000$	$n = 10000$
$\mathbb{P}(Y\|X = x)$	-0.296 +/- 0.090	-0.274 +/- 0.025	-0.277 +/- 0.007
Inc. bidder	-0.355 +/- 0.147	-0.294 +/- 0.034	-0.280 +/- 0.010

Table 2: Value of the *direct* incremental likelihood obtained thanks to the variable Y that we can observe on a simulated data set. We observe that without observing Y, the incremental factor achieves nearly as good performances as the theoretical maximizer $\mathbb{P}(Y\|X = x)$.

	$n = 100$	$n = 1000$	$n = 10000$
$\mathbb{P}(Y\|X = x)$	-0.289 +/- 0.016	-0.274 +/- 0.005	-0.275 +/- 0.002
Inc. bidder	-0.293 +/- 0.017	-0.280 +/- 0.006	-0.282 +/- 0.003

(a) observed clicks   (b) display type   (c) observed sales
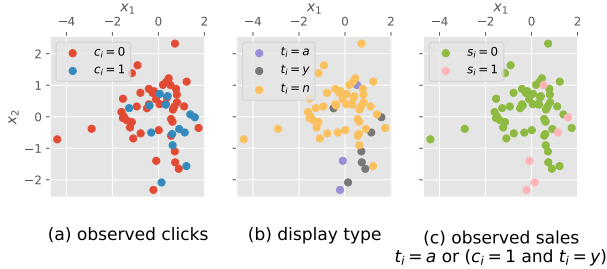$t_i = a$ or ($c_i = 1$ and $t_i = y$)

Figure 6: Illustration of the generative model described in Table 1 where each display is depicted by its feature vectors $x \in \mathbb{R}^2$. On a real data set we observe only if a display is clicked (a) and if it leads to a sale (c), but our variable of interest is the unobservable display type (b). Though, the metric we have introduced in Proposition 5.1 allows us to evaluate the capacity of a model to predict the unobservable event $T = y$.

Since the data is generated, we have access to the true value of the display type t_i and more specifically to the binary variable y_i (that we identify to the event $t_i = y$) that characterizes an incremental display. Hence, for a model $f : \mathbb{R}^2 \rightarrow (0, 1)$ supposed to predict y_i, we can compute the averaged direct incremental likelihood over n displays $\frac{1}{n} \sum_{i=1}^{n} \text{LLH}_B(y_i, f(x_i))$. We provide in Table 2, the results for two models. First, one that directly predicts the probability $\mathbb{P}(Y|X = x_i)$. Similarly to the metric, this model is trainable only if the label y_i is observed, i.e. only on a simulated data set. The second one actually predicts the incremental part of the bidding formula (3), with two models, one trained to predict $\mathbb{P}[S|C, X = x_i]$ and a second one to predict $\mathbb{P}[S|\overline{C}, X = x_i]$. This term is actually directly linked to $\mathbb{P}(Y|X = x_i)$ since from Equation (7), we derive

$$\mathbb{P}(Y|X) = \mathbb{P}[S|C, X = x_i] - \mathbb{P}(S|\overline{C}, X = x_i).$$

This observation makes our bidding formula a natural candidate to optimize for this incrementality metric and we indeed observe in Table 2 that its performances are really close to the one obtained with the theoretical model $\mathbb{P}(Y|X = x_i)$.

Finally, we present in Table 3 the value we obtain by using the reverted incremental likelihood from Proposition 5.1: $\frac{1}{n} \sum_{i=1}^{n} \text{RLLH}_B(c_i, s_i, f(x_i))$ given the predictions $\mathbb{P}(C|X = x_i)$ for all $i = 1, \ldots, n$ obtained from another model trained separately. We numerically illustrate the unbiasedness of the reverted incremental likelihood since we observe that it converges to the same values as the direct one. While it has more variance (leading to wider 95% confidence intervals), it is interesting to emphasize that this metric does not need to observe the hidden variable Y to correctly converge.