Optimal Bidding Strategies in Online Display Advertising

Prior Solutions

March 7, 2018

Abstract

Contents

1 Introduction

In these notes I describe a novel bid optimization strategy for display advertising campaigns based on real-time bidding (RTB) (see appendix A). When a campaign is launched, the advertiser uploads ad creatives, defines target rules (such as user segmentation[1], time and location) and the lifetime budget. Our goal here is to build an optimal bidding engine that maximizes, for a given budget, some chosen key performance indicator (KPI), such as the number of conversions or clicks, for a given budget.

2 Generalities

3 Outline of the problem

The bidding strategy is a function that maps the *individual impression evaluation* to a bid value:

$$b : \text{individual impression evaluation} \rightarrow \text{bid value} \tag{1}$$

Examples of individual impression evaluations are the click-through rate (CTR) and the conversion rate (CVR) which are defined as:

$$\text{CTR} = \frac{\#\text{clicks}}{\#\text{impressions}} \times 1000 \tag{2}$$

$$\text{CVR} = \frac{\#\text{conversions}}{\#\text{clicks}} \times 1000. \tag{3}$$

Note that the CVR in the data is the conversion rate *per click* and not per impression (which is more common). The campaign's individual impression evaluation is chosen by the advertisers. Both CTR and CVR are estimated *in real time* based on the impression features (see Fig. 6). The bid request corresponding to some impression can be represented by a high-dimensional vector $\mathbf{x}$. Examples of entries are:

- Auction features such as Ad slot ID which shows the location where the impresssion will show up

- Ad features such as creative ID

[1]Examples of segments are shoes & bags, travelling, clothing and others.

- Region ID
- Paying price a.k.a. *market price*, which is the auction winning price i.e. is the highest bid from the competitors and it is what the advertiser will pay for the impression
- User feedback on the ad impressions (number of clicks and conversions)

The vector $\mathbf{x}$ is then:

$$\mathbf{x} = (\text{Log type : Ad slot ID : } \ \text{Paying price ID : Region ID ...})^T \tag{4}$$

The output is the bidding price $b(k(\mathbf{x}))$ where the argument $k(\mathbf{x})$ is some predicted key performance indicator (KPI) chosen by the advertiser, such as CTR and CVR. We

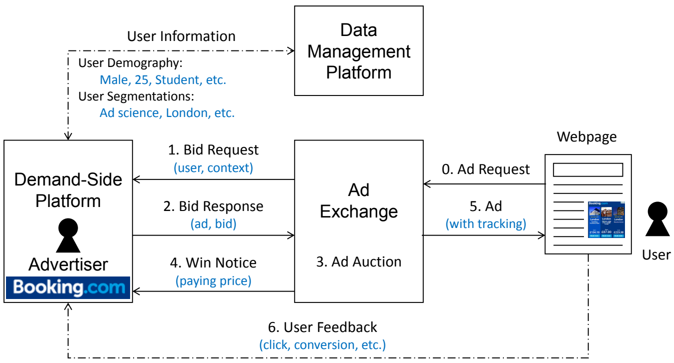

Figure 1: Real-time bidding (RTB) mechanism for display advertising (simplified).

.

denote by $p_{\mathbf{x}}(\mathbf{x})$ the prior distribution of the feature vectors matching the campaign target rules. For each campaign, the historic bidding and feedback data (CTR or CVR) are used by the advertiser to predict the KPI for the ad impression being auctioned. We denote the predicted KPI of a bid request $\mathbf{x}$ as $k(\mathbf{x})$ (different advertisers might consider different KPIs). If for example, the goal of a campaign is to maximise the total number of clicks, then $k(\mathbf{x})$ is the predicted CTR (pCTR) for that impression. If, instead, the goal is the maximization of the total number of conversions, then $k(\mathbf{x})$ is the pCVR for that impression.

It is shown here that the optimal bidding function is a non-linear, concave, function of impression evaluation metrics, such as CTR and CVR. As we shall see, in comparison with the linear bidding function/strategy that is widely used in the industry nowadays, this non-linear concave bid function $b(k(\mathbf{x}))$ leads to much better results and to novel (in relation to the linear strategy) recommendations for buying impressions. The function $b(k(\mathbf{x}))$ will be obtained from the analysis of the winning probability, or win-rate $w(b(k(\mathbf{x})), \mathbf{x})$.

The procedure is schematically as follows. Let's use conversions and the conversion rate (CVR) to exemplify: (1) We receive a bid request which is a log with features such as the one shown in Figure 6, represented by a vector $\mathbf{x}_0$, (2) we then use the training dataset to build a predictive model for the CVR and then use it to calculate pCVR for that particular impression corresponding to that particular bid request log $\mathbf{x}_0$, (3) the pCVR is then plugged into the bidding engine and the output is the bid price associated with that pCVR. In section 5 it will be argued that the arguments of the win-rate function and the bidding function can be simplified to:

$$w(b(k(\mathbf{x})), \mathbf{x}) \to w(b), \quad b(k(\mathbf{x}), \mathbf{x}) \to b(k) \tag{5}$$

It must be noted that, though the argument of the bidding function will be just the predicted KPI k, in this case pCVR,

$$b = b(p\text{CVR}) \tag{6}$$

and therefore b does not depend explicitly on the features vector, the other factors such as the budget, winning rate, market price and prior distribution, are implicitly embodied in the functional form of the bidding function since function $b(k)$ is obtained using a functional optimization technique which takes into account all those factors.

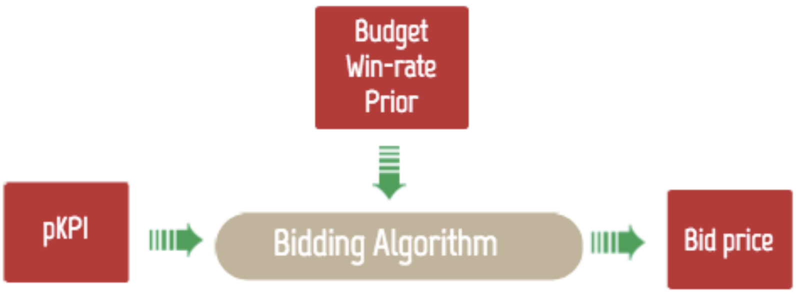

Figure 2: Inputs and output of the bidding engine.

Schematically, the problem is to find

$$\max_{b(\dots)} k_{\text{target}}(b(pk)) \text{ subject to cost} \leq B, \tag{7}$$

where $b(\dots)$ is the optimal bidding function, k_{target} is some chosen KPI to be maximized and pk is some predicted KPI rate corresponding to k_{target} (e.g. if the latter is the number of conversions, the former is the predicted conversion rate pCVR) and B is the campaign's budget. This is a *functional* optimization problem. A few examples are:

- Maximize the number of clicks

- Maximize the number of conversions
- Maximize a combination of clicks and conversions. In fact, in the iPinYou competition [3], the KPI was the following linear combination:

$$\#\text{clicks} + N \times \#\text{conversions}$$

where the parameter N was included to show the relative importance of clicks and conversions.

To predict the CTR or CVR one can use machine learning algorithms applied to the training dataset.

4 A simpler strategy as warm-up

I will briefly describe the linear bidding strategy since it is the most used strategy in the market nowadays (the same strategy is used throughout the whole lifetime of the campaigns). This strategy is based on a celebrated theorem from game theory according to which, for strategic competitors, the dominant strategy for pure Vickrey or second-price auctions is *truth-telling*, which means that bidders will *bid their private values*.

Suppose the advertiser's valuation of a conversion (or click) is equal to some target $k_{\text{target}}(\mathbf{x})$. As a participant in a real-time auction, the advertiser subscribes to a stream

Training Data

Adv.	Period	Bids	Imps	Clicks	Convs	Cost	Win Ratio	CTR	CVR	CPM	eCPC
1458	6-12 Jun.	14,701,496	3,083,056	2,454	1	212,400	20.97%	0.080%	0.041%	68.89	86.55
2259	19-22 Oct.	2,987,731	835,556	280	89	77,754	27.97%	0.034%	31.786%	93.06	277.70
2261	24-27 Oct.	2,159,708	687,617	207	0	61,610	31.84%	0.030%	0.000%	89.60	297.64
2821	21-23 Oct.	5,292,053	1,322,561	843	450	118,082	24.99%	0.064%	53.381%	89.28	140.07
2997	23-26 Oct.	1,017,927	312,437	1,386	0	19,689	30.69%	0.444%	0.000%	63.02	14.21
3358	6-12 Jun.	3,751,016	1,742,104	1,358	369	160,943	46.44%	0.078%	27.172%	92.38	118.51
3386	6-12 Jun.	14,091,931	2,847,802	2,076	0	219,066	20.21%	0.073%	0.000%	76.92	105.52
3427	6-12 Jun.	14,032,619	2,593,765	1,926	0	210,239	18.48%	0.074%	0.000%	81.06	109.16
3476	6-12 Jun.	6,712,268	1,970,360	1,027	26	156,088	29.35%	0.052%	2.532%	79.22	151.98
Total	-	64,746,749	15,395,258	11,557	935	1,235,875	23.78%	0.075%	8.090%	80.28	106.94

Test Data

Adv.	Period	Imps	Clicks	Convs	Cost	CTR	CVR	CPM	eCPC	N
1458	13-15 Jun.	614,638	543	0	45,216	0.088%	0.000%	73.57	83.27	0
2259	22-25 Oct.	417,197	131	32	43,497	0.031%	24.427%	104.26	332.04	1
2261	27-28 Oct.	343,862	97	0	28,795	0.028%	0.000%	83.74	296.87	0
2821	23-26 Oct.	661,964	394	217	68,257	0.060%	55.076%	103.11	173.24	1
2997	26-27 Oct.	156,063	533	0	8,617	0.342%	0.000%	55.22	16.17	0
3358	13-15 Jun.	300,928	339	58	34,159	0.113%	17.109%	113.51	100.77	2
3386	13-15 Jun.	545,421	496	0	45,715	0.091%	0.000%	83.82	92.17	0
3427	13-15 Jun.	536,795	395	0	46,356	0.074%	0.000%	86.36	117.36	0
3476	13-15 Jun.	523,848	302	11	43,627	0.058%	3.642%	83.28	144.46	10
Total	-	4,100,716	3,230	318	364,243	0.079%	9.845%	88.82	112.77	-

Figure 3: Dataset statistics of iPinYou competition for the impression/click/conversion logs

.

of bid requests, each of them representating an opportunity to show an ad to some particular user. The expected value of the click or conversion (supposing the advertiser won the auction) is simply:

$$v_{\text{lin}}(k(\mathbf{x})) = k(\mathbf{x}) \times k_{\text{target}}(\mathbf{x}) \tag{8}$$

where $k(\mathbf{x})$ is the predicted KPI. If the bidding is truthful $b_{\text{lin}} = v_{\text{lin}}$ and therefore:

$$b_{\text{lin}}(k(\mathbf{x})) = k(\mathbf{x}) \times k_{\text{target}}(\mathbf{x}) \tag{9}$$

It should be noted that the use of this theorem changes the ill-defined problem of finding an optimal bidding price into a well-defined problem, which is to predict the value of some KPI i.e. to find $k(\mathbf{x})$. If the KPI is the number of clicks, the problem is well-known and can be solved using a large variety of models (see next section). However, if the KPI is the number of conversions, there are issues with the sparcity of the data and one has to model the conversions at different hierarchical methods, as will be discussed later (see appendix E).

5 Non-linear bidding strategy

This *rationale* above, however, is not valid if the budget is fixed or if the advertiser is participating in more than one second-price auction. In Ref. [9] the competitive landscape in Ad Exchanges is studied. It is shown that, in repeated budget-constrained auction games, i.e. when advertisers have budget limits and participe in multiple Vickrey auctions over the lifetime of the campaign, dynamic interactions among them must be taken into account when the bidding landscape is characterized and truth-telling is not the a dominant strategy anymore.

Though the model used in Ref. [9] leads to useful insights, the approach taken here is more pragmatic and uses simple statistical methods to model the market price and winning probability, instead of assuming the advertisers have well-defined strategies and will bid their private value.

In this analysis, the bid price b will depend on many other practical factors, and not only on the predicted KPI value of the impression being auctioned. Among those factors are the budget constraint, the auction winning probability (or win-rate), the market price distribution, the feature vector of this particular impression and the prior distribution of x. The problem then becomes significantly more convoluted (as will be discussed later, for very weak budget constraints, the theorem about Vickrey auctions can be used and the corresponding linear strategy works reasonably well).

Market prices are modelled as a stochastic variable since it is virtually impossible to analyse the bidding strategy of all auction participators. To predict them we can either use the data statistics or a regression model with censored data. The first option is more straighforward and in simpler to use in practice. The second option however is more thorough and will be discussed in appendix B where it is heuristically derived an algorithm that learns the distribution of winning bids based on partial information (such as this case) and suggests bidding actions to maximize the number of impressions given a budget constraint.

By analytically solving the optimization problem, it will shown that the win-rate function $w(b)$ is much more important than the distribution of the features vector $\mathbf{x}$ in the process of building the optimal bidding function b (the $\mathbf{x}$ probability distribution is found to be only weakly correlated with the form of b).

Before deriving the new bidding function taking into consideration this information, we need to obtain a mathematical expression for the winning-rate. Following [2] we will derive an expression for the winning rate using functional optimization. The predicted KPI will be denoted by $k(\mathbf{x})$. For example, if the campaign goal is to maximise conversions, $k(\mathbf{x})$ denotes the predicted conversion rate $p\mathrm{CVR}$ for that impression. We also denote by $p_k(k)$ the prior distribution of the predicted KPI per bid request.

Note that when the campaign starts there is no available data. To deal with this issue, one possibility is to spend a small fraction of the budget, bid randomly and learn some statistics. In [5] for example, a bid landscape prediction, i.e. the auction volume forecast, is used to estimate the auction statistics for a given setting and budget constraint. The estimated number of bid requests for the target rules during the lifetime T is denoted by $N(T)$. Hence we see that the approach has two stages, the first one is learning the statistics to obtain $p_{\mathbf{x}}(\mathbf{x})$ and $N(T)$, and the second is to find the functional form for the winning rate and for the corresponding bid.

To derive a functional form for the winning rate $w(b(k(\mathbf{x})))$ let us formulate the problem more precisely.

Let us suppose the goal of the campaign is to maximize some specific KPI, say, the number of clicks or conversions. Now, for concreteness, let's consider CVR to be the quantity inside the argument of the bid function a write $k(\mathbf{x}) = n(\mathbf{x})$. The functional optimization problem can then be stated as:

$$
\begin{aligned}
b(n(\mathbf{x})) = \underset{b(n(\mathbf{x}))}{\arg\max}\, N(T) \int d\mathbf{x}\, n(\mathbf{x}) w(b(n(\mathbf{x}))) p_{\mathbf{x}}(\mathbf{x}) \\
\text{subject to} \quad N(T) \int d\mathbf{x}\, b(\mathbf{x}) w(b(n(\mathbf{x}))) p_{\mathbf{x}}(\mathbf{x}) \leq B
\end{aligned} \tag{10}
$$

5.0.1 Remarks

A few remarks are in order:

1. The product of the first two factors inside the integrand of $b(n(\mathbf{x}))$ is:

$$
\begin{aligned}
&n(\mathbf{x})w(b(\mathbf{x})) \\
&\quad = \text{conversion rate} \times \Pr(\text{winning auction})
\end{aligned}
\tag{11}
$$

   which is the probability of conversion provided the auction was won. This is for a given feature vector $\mathbf{x}$. The integral sum over feature vectors. The integral is then multiplied by the number of auctions $N(T)$ during the interval T.

2. This is a *functional* optimization problem i.e. we are after a functional $w(b(\mathbf{x})) = f(b(\mathbf{x}))$ or "a function of a function" of $\mathbf{x}$ and $\mathbf{x}$ is not explicitly present in $w(b(\mathbf{x}))$.

3. The function $p_{\mathbf{x}}(\mathbf{x})$ is the probability distribution of the feature vectors $\mathbf{x}$. It is the probability (density) that the bid request will be, say, for an impression placed in the Desktop news feed, on a Monday, at 11pm, etc. Naturally, this will be available only after the learning phase.

4. The product $b(\mathbf{x})w(\mathbf{x})$ inside the integrand on the second line of Eq. (10) is the expected value of the bid for a given feature vector $\mathbf{x}$. Integrating over $\mathbf{x}$ and multiplying by the number of bid requests, we sum over all bids, and this sum must be restricted by some budget B. Note that, due to the reserve price setting, the cost is usually higher than the second highest bid and we use the bid price $b(\mathbf{x})w(\mathbf{x})$ as the upper bound of the cost of winning.

5. We note that, as stated, the problem is not explicitly dynamical, but the time-dependency is included by considering date and time as features.

6. Notice that by simultaneously maximizing the number of clicks and fixing a budget, *we are indirectly minimizing the predicted cost of each click or conversion i.e.* CPC *or* CPA . These are extremely important metrics in display advertising.

7. The budget constraint and the campaign's lifetime are taken into account. The former by the condition to which the optimization is subject and the latter by the factor $N(T)$.

5.1 How to simplify the problem

The problem can be simplified by writing the probability distributions of $\mathbf{x}$ in terms of the prior probability distribution of $n(\mathbf{x})$ which we will denote by $p_{\mathbf{n}}(\mathbf{n})$. Their ratio is the gradient $\nabla n(\mathbf{x})$ and changing variables inside the integral the optimization problem reads:

$$b(n) = \arg\max_{b(n)} N(T) \int dn\, n w(b(n)) p_n(n)$$
$$\text{subject to} \quad N(T) \int dn\, b(n) w(b(n)) p_n(n) \leq B \tag{12}$$

The problem can be solved by writing down a Langrangian and using basic calculus of variations. The constrained Lagrangian would read:

$$L(b(n), \lambda) = \int dn\, n w(b(n)) p_n(n) - \lambda \left[\int dn\, b(n) w(b(n)) p_n(n) - \frac{B}{N(T)} \right]$$

Optimization of L leads to canonical Euler-Lagrange differential equations which applied to L gives:

$$n p_n(n) \partial_{b(n)} w(b(n)) - \lambda p_n(n) \left[w(b(n)) + b(n) \partial_{b(n)} w(b(n)) \right] = 0 \tag{13}$$
$$\lambda w(b(n)) - [n - \lambda b(n)]\, \partial_{b(n)} w(b(n)) = 0 \tag{14}$$

This system has no unique solution. A quite general form for $w(b)$ is:

$$w(b) = \frac{b^\alpha(n)}{c^\alpha + b^\alpha(n)} \tag{15}$$

where $\alpha \leq 2$ for analytical solutions. Analyzing the dependency of the winning bid prices (the market prices) on the bid request features [2] , one finds that that the winning price distributions have no clear relation to the feature values. This is consistent with this expression for $w(b)$, according to which the bid price is the only argument i.e. we can write $w(b, \mathbf{x}) \equiv w(b)$. In [2] the values of $\alpha = 1, 2$ are chosen where c is a constant. Using $\alpha = 1$ we obtain the following expression for the bidding function:

$$b(n) = c \left[\sqrt{\frac{1}{c\lambda} n + 1} - 1 \right] \tag{16}$$

where k can the estimated CTR or estimated CVR. This expression is clearly concave. The open parameter c is obtained from the dataset. To find λ, $b(k)$ must be substituted into the Lagrangian constraint (where the inequality would become an equality):

$$\int dk\, b(k, \lambda) w(b(k, \lambda)) p_k(k) = \frac{B}{N(T)}, \tag{17}$$

which states that all budget was used. The variable λ was included for convenience (see below). This equation can be solved numerically. In [2] however, they suggest a different approach which takes λ as a tuning parameter based on the training dataset. From equation (16) we see that, as λ decreases, $b(k)$ increases. The win-rate function

$$w(k) = 1 - \left(\frac{k}{c\lambda} + 1 \right)^{-1/2}, \tag{18}$$

It is not difficult to check that $b(k)w(b(k))$ increase monotonically as λ decreases. The same behavior occurs for the integral since $p_k(k)$ is independent of λ. Hence, increasing the budget per-case $\frac{B}{N(T)}$ implies that λ decreases and that corresponds to a higher bid price. The optimal λ trend as a function of $\frac{B}{N(T)}$ is shown in Figure 5.

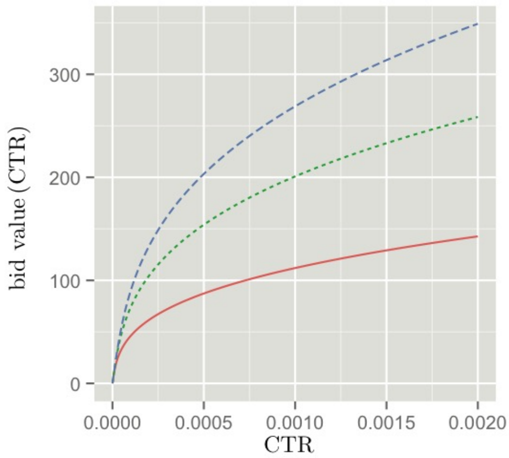

Figure 4: Example of bidding functions where the argument is the CTR.

.

5.2 Tuning λ

6 Data

6.1 Description of the Logs

There are four types of logs in the training dataset:

- Bids logs
- Impression logs
- Click logs
- Conversion logs

Examples are shown in Figure 6. The log on the left-hand side is the bid log, where the first entry, Bid ID, uniquely identifies one ad impression. Corresponding to this bid log there are two other logs, namely, one impression log and one click log (or conversion log), both with the same Bid ID. Hence, the Bid ID can be used to join the bidding, impression and click logs.

The 24 features in the dataset are shown in Figure 6. The dataset is from a well-known DSP company and contains more than 15 million impressions with user feedback of 9 campaigns from multiples advertisers in an interval of 10 days in 2013. The logs are on a row-per-record basis. There are two types of logs, bid logs and impression/click and conversion logs

Figure 5: Tuning of the λ parameter.

.

For each log there is information about the user, about the advertisers (characteristics of the creative such as format and size, etc), about the publisher (the auction reserve price, page domain, URL and ad slot, etc). For each bid request the user clicks/conversions is recorded if the advertiser won the auction.

More specifically, for each record (row), some of the following types of entries/-columns [3](for more details see here) are present:

- Auction features (columns 1, 2, 4-12)
  - The bid ID identifies the event log. Note that a record (row) contains not only information about clicks/conversions. Part of the logs correspond to impressions.
  - The date format is yyyyyMMddHHmmssSSS
  - The Log type is either 1, 2 or 3 corresponding respectively to an impression, a click or a conversion.
  - The User-Agent column describes the device, OS and browser of the user.
  - Columns 10 and 11 correspond to the hosting webpage of the ad slot being auctioned, and the correspondent URL.
  - Column 18 define a minimum bid to win the auction.

SN	Column	Example
*1	Bid ID	c0550000008e5a94ac18823d6f275121
2	Timestamp	20130218134701883
*3	iPinYou ID	35605620124122340227135
4	User-Agent	Mozilla/5.0 (Windows NT 5.1) \ AppleWebKit/535.11 (KHTML, \ like Gecko) Chrome/17.0.963.84 \ Safari/535.11 SE 2.X MetaSr 1.0
*5	IP	119.163.222.*
6	Region ID	146
7	City ID	147
8	Ad Exchange	2
*9	Domain	e80f4ec7f5bfbc9ca416a8c01cd1a049
*10	URL	hz55b000008e5a94ac18823d6f275121
11	Anonymous URL	null
12	Ad Slot ID	973726_9023493
13	Ad Slot Width	300
14	Ad Slot Height	250
15	Ad Slot Visibility	FirstView
16	Ad Slot Format	Na
17	Ad Slot Floor Price	0
18	Creative ID	f80f4ec7f5bfbc9ca416a8c01cd1a049
*19	Bidding Price	573
20	Advertiser ID	2259
*21	User Profile IDs	null

SN	Column	Example
*1	Bid ID	01530000008a77e7ac18823f5a4f5121
2	Timestamp	20130218134701883
3	Log Type	1, 2, 3
*4	iPinYou ID	35605620124122340227135
5	User-Agent	Mozilla/5.0 (compatible; MSIE 9.0; \ Windows NT 6.1; WOW64; Trident/5.0)
*6	IP	118.81.189.*
7	Region ID	15
8	City ID	16
9	Ad Exchange	2
*10	Domain	e80f4ec7f5bfbc9ca416a8c01cd1a049
*11	URL	hz55b000008e5a94ac18823d6f275121
12	Anonymous URL	null
13	Ad Slot ID	2147689_8764813
14	Ad Slot Width	300
15	Ad Slot Height	250
16	Ad Slot Visibility	SecondView
17	Ad Slot Format	Fixed
18	Ad Slot Floor Price	0
19	Creative ID	e39e178ffdf366606f8cab791ee56bcd
*20	Bidding Price	753
*21	Paying Price	15
*22	Landing Page URL	a8be178ffdf366606f8cab791ee56bcd
23	Advertiser ID	3358
*24	User Profile IDs	123,5678,3456

Figure 6: The four types of logs in the training set. The column on the right represents three columns one for each log type.

.

  - Auction winning price (column 21) is the highest bid from competitors (since we have only second-price auctions). If the bidding engine gives a bidding price higher that the auction winning price, the advertiser gets the ad impression corresponding to that Bid ID, and this record will occur in the impression log. The distribution of the market price may be estimated examining the data statistics or using a regression model, which is called bid landscape forecasting. The knowledge of the market price distribution allows one to estimate the probability of winning a given ad auction and the corresponding cost given a bid price.

- Ad features (columns 13-19 and 22-24)
  - Column 16 describes how fast the ad is viewed depending on its position. For example, if the ad slot is above the fold it is classified as “FirstView". If one has to “scroll down” to see the ad, it is classified as “SecondView” and so on.
  - Column 17 can be “Fixed” (fixed size and position), “Pop” (pop-up), etc

- User feedback, measured by clicks or conversions, on that impression (column 3). The “User Tags” column will tell us the segments if which the user in interested in e.g. buying something, travel, readings, etc.

The auction and ad features are sent to the bidding engine to generate a bid response. Note that log type, paying price and the key page URL are present only in the impression/click/conversion logs (and not in the bid logs).

7 Experiment

The dataset was described in section 6. We will briefly review it here. Consider Figure 6. Note that the log on the left hand side, called the bid log, does not have information about Log type, paying price and Key Page URL. The Log type entry, present in the three logs on the right hand side determines if the log corresponds to an impression, a click or an impression. The auction winning price, or paying price or market price in column 21 is the highest bid from the competitors. The logs contain information about:

- The user. For example User tags determines user segmentation (if the user is interested in travelling, reading, etc)
- The advertiser (e.g. format and size of creative)
- Publisher (e.g. ad slot size)

In Figure 8, the relation between the auction winning rate and the bid value from the training dataset is shown. Eq. (16) is therefore a reasonable form for the $w(b)$. The dependence of the winning bid price (column 21) is shown in **??** not to depend strongly on the features values (this is operationally obtained by comparing, for each impression log (Log price $= 1$) the relation between the value on column 21 and the feature entries on that log). This suggests that the features are not very important in the determination of the winning rate which corroborates the ansatz $w(b(\mathbf{x}), \mathbf{x}) = w(b(\mathbf{x}))$.

Since, according to these curves, the impressions with low cost (low bid value) are more cost effective, this bidding strategy encourages the advertiser to bid more low cost impressions. In other words, the bidding strategy should try to bid more impressions instead of focusing only on the small set of impressions with high value.

8 Training/testing

8.1 Training

The data is split with ration 2:1 by Timestamp sequence. The training data is used to:

- Train the predicted CTR or CVR estimator extracting the feature values from the log data. Machine learning models are used here.
- Tune the parameters c and λ of the bidding function.

8.2 Testing

The test data is a list records consisting of the features of a bid request, the auction winning price and user feedback information (CTR or CVR for example). The bidding engine reads each record and generates a bid price. The bid price is compared with the auction winning price and in case it is higher the ad is shown and the clicks/conversions and charged price of that record are used to update the performance and the total cost. In more details we have the following steps:

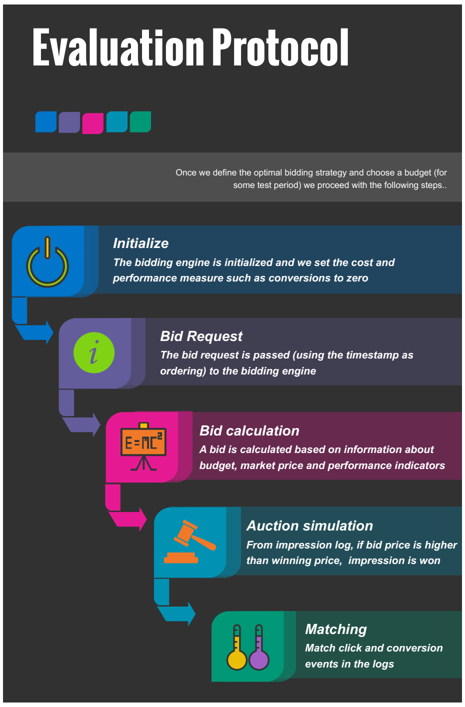

Figure 7: Evaluation protocol.

1. First initialise the bidding engine, setting a predefined budget, initialising the cost and performance as zero. Measures of performance are, for example, achieved clicks and achieved number of conversions.

2. A bid request is passed (ordered by the timestamp feature) to the bidding engine. It contains contextual and behavioural data of the corresponding auction and also the ad data as shown Figure 6.

3. With information such as the budget, current cost and achieved performance, the bidding engine computes a bid for this request. If at some point the cost surpasses the budget, all left bid requests are skipped (the bid responses are set to zero after this point).

4. The auction is simulated by referencing the impression logs (see Fig. 6 on the right) if bid price is higher than the auction winning price in the log (paying price on column 21) and floor price (column 18), the bidding engine wins the auction and gets the ad impression.

5. The next step is to match the click and conversion events in the logs for this impression if the auction is won. The performance of the bidding machine is updated and saved and the cost increases by the paying price.

6. Check for any bid requests left in the test data and terminate the process if there aren't any.

8.3 Impact of budget

The budget was chosen to be equal to $1/64, 1/32, 1/16, 1/8, 1/4$ and $1/2$ of the original total cost of the test data. It is found that when the budget is small, the non-linear bidding strategy improves the results (number of clicks or conversions) significantly. If the budget is less restricted, simpler strategies also deliver reasonably good results. Nevertheless, non-linearity is found to be optimum in all scenarios. In Figure 10 we see the comparison between the strategy currently used in the market and the strategy explained here. In [2] it is shown that the non-linear strategy outperforms all others both in number of clicks and eCPC.

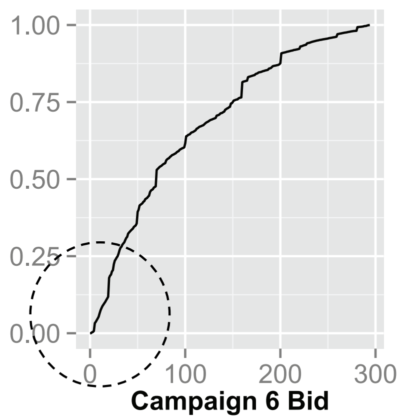

Figure 8: Win-rate curve for one of the campaigns. We see that lower evaluated impressions are more cost effective.

.

A Real-time bidding

In real-time bidding (instantaneous) auctions [1], advertising inventory is bought and sold on a per-impression basis. Advertisers bid on individual impressions and it the bid is won, the corresponding ad is displayed instantly on the publisher's website. Typically the process begins when the user visits a website, triggering a bid request to an

Bidding Strategies	Const	Rand	Mcpc	Lin	ORTB
Consider budget conditions	√	√		√	√
Evaluate per impression value			√	√	√
Consider winning function					√

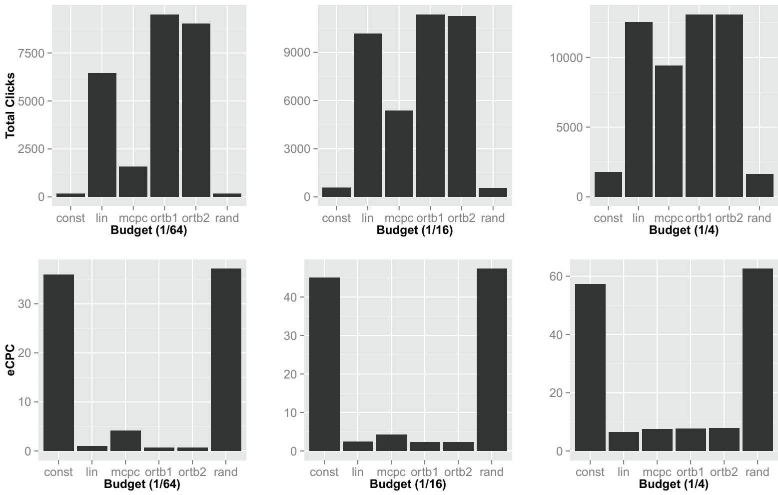

Figure 9: Strategies comparison. Only the linear and ORTBs are relevant in this discussion.

.

ad exchange. The bid request contains data such as user's demographics, browsing history, location, etc. Ad exchanges submit the request to advertisers who automatically bid, in real time, to place their ads (this is done on each impression as it is served and it is repeated for every ad slot on the webpage). The highest bidder wins the impression and their ad is served on the webpage.

B Predicting Winning Price in RTB with Censored Data

Consider the linear regression

$$z = \beta^T \mathbf{x} + \epsilon \tag{19}$$

where the noise is Gaussian. The winning price is only observed when the bid wins the auction which implies that the instances of $(\mathbf{x}, z)$ observed a *right-censored* and biased. In [12] a censored linear regression is used to model the winning price as a

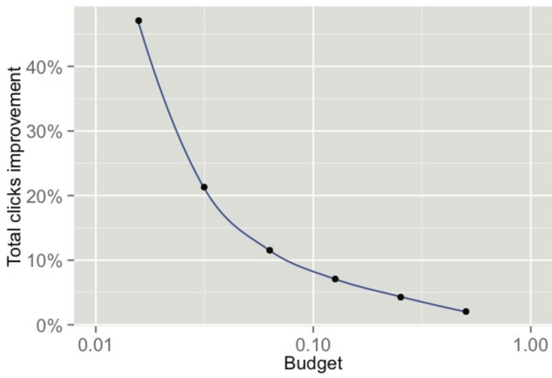

Figure 10: The number of clicks improves using the strategy described here (compared with the linear strategy).

function of the auction features. The distribution of the winning price is

$$p(z) = \log \phi \left(\frac{z - \beta^T \mathbf{x}}{\sigma} \right) \tag{20}$$

where ϕ is the standard normal density function. For the auctions lost the data $(\mathbf{x}, b)$, where b is the bid price the only information we have is that winning price is higher than the bid $z > b$ and $P(z > b)$ is defined as:

$$P(z > b) = \log \Phi \left(\frac{\beta^T x - b}{\sigma} \right) \tag{21}$$

is the cummulative stardard normal. We conclude that if the observed data $W = (\mathbf{x}, z)$ and censored data $L = \mathbf{x}, b$ we have the following training of the censored linear regression:

$$\min_{\beta} \; - \sum_{(\mathbf{x},z) \in W} \log \phi \left(\frac{z - \beta^T \mathbf{x}}{\sigma} \right) - \sum_{(\mathbf{x},b) \in L} \log \Phi \left(\frac{\beta^T \mathbf{x} - b}{\sigma} \right) \tag{22}$$

C Click-through rate model

D Bid Landscape Forecasting

E Conversions and the problem with sparcity

References

[1] Real-time bidding, Wikipedia contributors, Wikipedia, 2 June 2017, The Free Encyclopedia.

[2] Zhang et al, Optimal Real-Time Bidding for Display Advertising.

[3] Real-Time Bidding Benchmarking with iPinYou Dataset, Weinan Zhang, Shuai Yuan, Jun Wang, Xuehua Shen, arXiv:1407.7073

[4] Feedback Control of Real-Time Display Advertising

[5] Y. Cui, R. Zhang, W. Li, and J. Mao. Bid landscape forecasting in online ad exchange marketplace. In KDD, pages 265–273. ACM, 2011

[6] Real-Time Bidding Benchmarking with iPinYou Dataset

[7] Statistical Arbitrage Mining for Display Advertising

[8] More precisely, the winner pays “one cent” more than the second higher bid.

[9] Balseiro, S. R., Besbes, O., and Weintraub, G. Y. (2015). Repeated auctions with budgets in ad exchanges: Approximations and design. Management Science, 61(4): 864–884.

[10] Graepel et al, Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft’s Bing Search Engine

[11] Yuan S. and Wang J., Real-Time Bidding: A New Frontier of Computational Advertising Research, A WSDM 2015 Tutorial

[12] Wush Chi-Hsuan Wu, Mi-Yen Yeh, Ming-Syan Chen, Predicting Winning Price in Real Time Bidding with Censored Data, KDD 2015.

[13] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft’s bing search engine.

[14] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads.