2014 International Conference on Management Science & Engineering (21th)
August 17-19, 2014 Helsinki, Finland

Optimized Bidding Algorithm of Real Time Bidding in Online Ads Auction

ZHANG Chong-rui, ZHANG E
School of Information Management and Engineering,
Shanghai University of Finance and Economics, Shanghai 200433, P.R.China

Abstract: RTB is one of the most popular topics on online advertising in recent years; RTB-based ads auction subverts traditional advertising strategy, explicitly stating the concept that "advertisement is commodity". Most of the current researches on online ads auction focus at the point of linear programming or assignment problem, while there's little study on the macro-level auction mechanism, especially the characteristics of the general winning bids. Based on real log data from a DSP in China, this paper conducts empirical analysis of average deal price and bidding volume at different times of impressions. Finally, we design an optimized method for delivery of RTB ads through offline experiments.

Keywords: auction mechanism, empirical research, online advertising, real-time bidding

1 Introduction

RTB, namely "Real Time Bidding", is a way to sell advertising on the network by targeting right audience and assigning each impression to the viewer **[1]**. Currently, the analysis and processing of data generated for RTB is focusing on micro-level "audience targeting" and group labeling technology. As for macro-level auction markets, as well as the relation of commodity supply and demand, these perspectives have not gained much attention. The goal of this paper is to study the characteristics of general RTB auction marketplace, especially the relationship between deal volume and price in different time periods, which is revealed by real log data from a Chinese demand-side platform and to design an experiment demonstrating our empirical analysis, therefore proposing one approach of real-time bidding algorithm optimization from auction mechanisms.

1.1 RTB ecosystem

RTB ecosystem is constituted of Advertisers who have demands to display their ads, Publishers who provide ad placements slots, ad auction trading platforms **[6]**, as well as different agents, intermediaries, data management platform and other participants. In the model we discussed, the players include:

(1)Advertisers: including agencies who need impressions to display their advertising creative. The highest bidder will win ad impressions and the goal is to reach a certain group of customers across web pages under budget constraints. [4]

(2)Ad Exchange: an open market matching impressions with ad slots through real-time bidding. Ad Exchange is similar to a stock market, providing the buyers (advertisers) with an auction platform **[4]**, assuming that exchanges do not own advertising slots or demands for impressions.

(3)Media or Publisher: portals, vertical sites, web unions, etc. who hold willingness to sell ad slots.

(4)Demand-side platform (DSP): serving for advertisers to provide cross-platform and cross-media advertising, to reach accurate audience though data integration and to optimize budget allocation under real-time monitoring.

1.2 RTB matching procedure

By the time a viewer opens a webpage that contains an ad slot for RTB advertising, the real-time bidding process starts immediately. The Ad Exchange gathers many multimedia impression opportunities and sends biding request to a number of DSPs. Biding request includes information about the user, advertising slot and impression environment. After receiving bidding request, DSPs will retrieve the advertiser's database, match their accumulated cookie data with the Ad Exchange (i.e. cookie mapping) **[10]**, and then decide according to their own algorithms whether to respond and the specific offers of price if bidding. Ad Exchange makes a final decision on all of the bids received or within predetermined period of time (100ms). The highest bidder will win the opportunity to show his ads, display the client's ad on the audience's browser. And the final payment to this impression is the second highest price on the market, rather than the winner's own global highest price. The entire process, from opening the page to final ads demonstration, needs to be completed within a very limited time.

Supported by the Ministry of Education of Humanities and Social Science project (12YJC630284)

978-1-4799-5376-9/14/$31.00 ©2014 IEEE

1.3 Second-price sealed-bid auction

Nowadays, mainstream ad exchanges are using second-price sealed-bid auction mechanism **[5]** (hereinafter referred to as "second price"). "Second price" is a bidding process that the highest bidder wins the impression but actual payment by the winner is the second highest price of all bids in the market **[7].** "Second price" is a fundamental mechanism in this article. Despite the fact that winner's highest price cannot reflect the market's overall value assessment, the second highest price, or the deal price is closer to the overall bidding level. If second price is approaching the first or highest price, it indicates fierce competition during auction because, in addition to the winner, there is presence of other bidders on the market who reckon that this impression enjoys a relatively higher value. If the first price and the second price are far away, indicating that, except the winner, other bidders executed a low-cost bid, which facilitates the winner to easily obtain that impression. Therefore, we believe that it is feasible to choose second auction price as agent of general demand.

2 Literature review

Google Inc.'s S. Muthukrishnan proposed some theoretical models and researches on Ad Exchange. In his paper "Ad Exchanges: Research Issues" **[5],** he outlined several issues about bidding mechanism from multiple angles of game theory and algorithms. This paper is the empirical analysis of raw log data from the perspective of macro-based auctions.

The paper of "Real-Time Bidding Algorithms for Performance-Based Display Ad Allocation" **[8]** by Ye Chen and other researchers from Microsoft Microsoft has designed a basic RTB algorithm by estimating CTR. They were inspired by duality theory of LP, proved the feasibility of offline auctions to simulate real-time cases, and further considered the problem of budget constraints and frequency control. The bidding program designed in this article has referred its basic RTB algorithm about impression flows, but we focus more about the general deals in different periods of time without predicting CTR. So there is essential difference between the two algorithms.

Microsoft researcher Daniel Gray Goldstein designed a clever experiment to explore the impact of time length on advertising effectiveness in his award-winning paper "Improving the effectiveness of time-based display advertising" **[3]**. Although his research is not directly linked with this paper, we are still inspired by the idea of experiments.

Finally, "Adaptive Bidding for Display Advertising" **[9]** initially proposed theories like Fully Observable Exchange and proposed a Learn-Then-Bid algorithm. The difference within this paper is that they are learning cumulative distribution of the highest bid from market, while we care more about average deal price public to us.

In summary, researches on raw RTB log datasets are not quite common, and study of the relationship between supply and demand in various times is mainly concentrated in traditional areas similar to stock markets. However, unlike stock markets who will be shutdown daily, RTB auction platform is a very special marketplace due to one of its characteristics of 24*7 continuous transactions. In this case, the impact of time on ads auctions by real-time-bidding is worth analyzing.

3 Empirical analysis of original datasets

3.1 Original server log datasets

The research data sources are provided by one Chinese DSP. The platform is the first who successfully access the entire domestic Ad Exchange platforms, being exposed to over billion impressions daily. We collect the platform's log datasets recording biddings (BID datasets), impressions (IMP datasets) and clicks (CLK datasets), spanning two months, a total of 14 days in a two-week cycle (March 11- March 17, 2013 and June 6 - June 12, 2013). The BID datasets record each bid request received. The average amount of bidding daily in March was 1.9 million and average 2.3 million records every day in June. The IMP datasets contain the impressions or successful winning bids, including actual deal price after auction. We sampled totally 9,262,763 impressions in March and 12,237,087 in June. The CLK datasets summarize the clicks generated by audiences after impression and there are 17,460 in total. Tab.1 are part of the log formats and examples that we are using in the raw datasets:

Tab.1 Treatment to IMP datasets

Field	Example	Explanation
Bid ID(*)	01530000008a 77e7ac18823f 5a4f5121	Auction ID after hash treatment , used to uniquely identify an impression;
Timestamp	20130218 001203600	Record occurrence time of impression;
Floor Price (RMB / CPM)	0	The reserved price publisher allows for ad slot (linear scale);
Bid Price (RMB / CPM)	753	Bidding returned by DSP (linear scale);
Deal Price (RMB / CPM)	15	Actual price a winner paid (linear scale);

The fields utilized in BID and CLK datasets are identical.

3.2 Data aggregate

Since the completion of a bid must be controlled within the millisecond level and each auction should be required to accommodate various conditions of users and pages, the deal price volatiles intensely. The time granularity that is too meticulous may not produce obvious conclusions when discussing macro auction markets. Due to our aim to analyze the relationship between deal volume and price in different time periods, selecting each hour as a statistical granularity is appropriate and meaningful. Based on this granularity, we set an hour as grouping variable and the deal price as statistical variable, therefore getting the hourly average deal price. Set the “hour” as grouping variable and counts the amount of transactions or impressions in each time period, thus getting the deal volume of each hour. Set “hour” as grouping variable in CLK datasets, getting numbers of clicks per hour, and calculate the number of times audience clicked divided by the total impressions per hour. The actual conversion rate is estimated by **[8]:**

$$CTR = P\,(click \mid t) = \frac{\text{number of click in t}}{\text{number of impressions in t}} \quad (1)$$

For 14 days and a total of 336 hours, the data is described as follows in Tab.2:

Tab.2 Descriptive statistics

	N	Mean	Std.Dev	Min	Max
Deal (mean)	336	75.74	8.4543	54.9	100.97
Deal Volume	336	63987.95	41490.78	5395	419310
# of Clicks	336	51.96	41.030	1	401
Floor Price	336	14.31	26.766	0	295

After data preprocessing, with 24 * 14 = 336 hours in total, the average deal price, deal volume and click-through-rate data are summarized in Tab.3:

Tab.3 Summary of average deal price, deal volume, number of clicks and CTR in 24*14 hours

ID	Day	Hour	Deal	Volume	Clicks	CTR
1	1	00:00-01:00	76.42	62249	53	0.00122
2	1	01:00-02:00	67.32	28988	16	0.00232
3	1	02:00-03:00	76.52	16065	13	0.00476
4	1	03:00-04:00	75.86	21905	14	0.00346
5	1	04:00-05:00	79.43	8075	11	0.00983
6	1	05:00-06:00	78.52	6491	5	0.01209
7	1	06:00-07:00	72.47	19062	15	0.00380
…	…	…	…	…	…	…
335	14	22:00-23:00	67.51	63143	40	0.00106
336	14	23:00-24:00	73.62	78333	74	0.00094

It should be noted that here we do not consider the budget constraints or the impression frequency control issue, because our study is targeting on the impression data from entire marketplace and these winning bids themselves should be constrained by various DSPs’ budgetary costs and frequency control.

3.3 Empirical analysis

(1) Analysis on consecutive average deal price, volume, click-through-rate per hour

Let us first establish the relationship between the average deal price, volume and click-through-rate in continuous time periods. After observing Fig.1, Fig.2 & Fig.3, we have found that the three curves are in accordance with sine waves roughly in successive hours. We can institutionally conclude that, although bidding opportunities and related impression environment are coming up randomly, there still exists a pattern beneath average deal price levels, deal volumes (or times of ad impressions) and CTR per hour in RTB auction market. The deal volume curve fluctuates more regularly, with significant peaks and troughs. By the time deal volume curve reaches its peak, CTR curve is also at a respectively high level. However, the average deal price curve is not always synchronous with them.

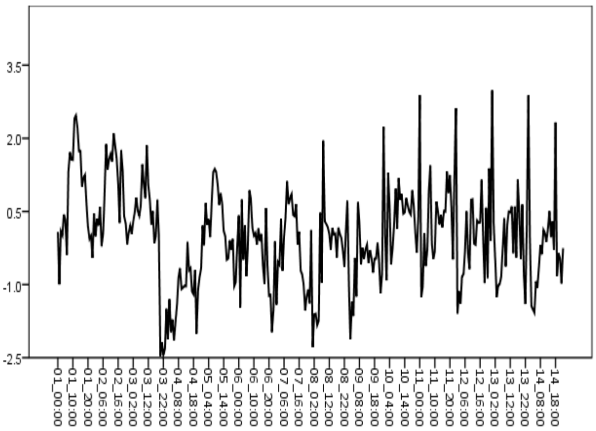

Fig.1 Continuous deal price curve

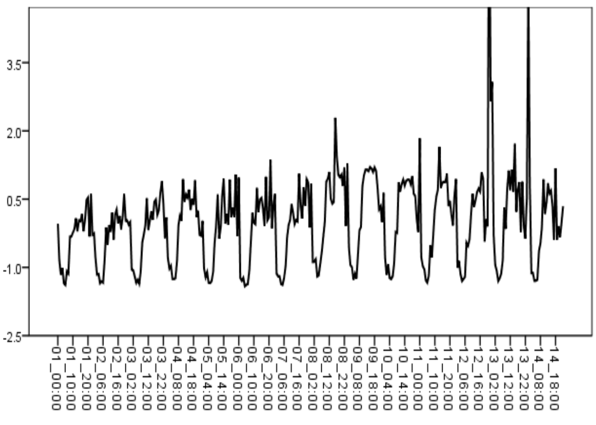

Fig.2 Continuous deal volume curve

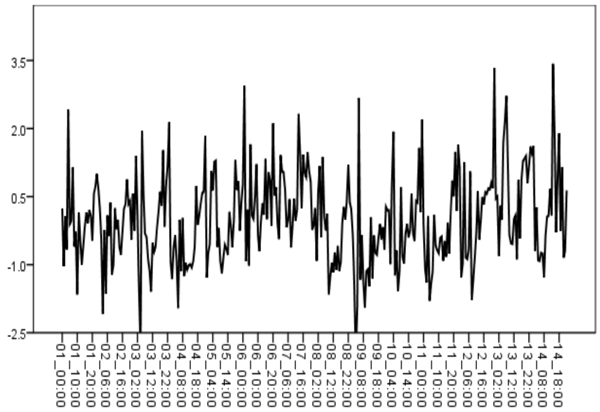

Fig.3 Continuous click-through-rate curve

*The consecutive lines represent average deal price per hour, deal volume per hour and CTR curve per hour; x-axis represents 24 * 14 hours' time points, y-axis is the statistical value; all data has been Z-SCORE normalized.*

The pattern beneath curve volatility reflects a pattern of universal web-browsing mode every day. When overall impression volume is large, the number of times audience would click will undoubtedly increase and deal volume is not stable during 24 hours. So in what specific time period will the click-through-rate become relatively high? When and what is the difference between average deal price and deal volume? When is the fiercest time of competition? The answers to these questions are quite crucial and helpful to understand the characteristics of RTB auction marketplace, but it is also very difficult to accurately grasp merely from consecutive time curve. Therefore, it requires that data from the 14-day-period should be compared within one figure. However, before conducting this, we must first verify that the average deal price & volume curve fluctuation at different time periods is real and easily proved.

(2) ANOVA on average deal price in different time sections

We have presented the Friedman test results for the 14-day average deal price data in Tab.4. The result shows that p-value <0.05, indicating a significant difference in deal price per hour. Tab.5 is the result of homogeneity of variance test on data introduced in Tab.3 using SPSS. The result shows that p-value <0.01, indicating significant differences in homogeneity of variance, and therefore we can continue ANOVA on average deal price **[12]**. Since we only need to figure out the pattern between average deal prices in time segments, not necessarily specific points of hours at beginning, so we initially divide 24 hours into four interval segmentations, 6 hours for each section: Dawn (00:00 to 06:00), Morning (06:00-12:00), Afternoon (12:00-18:00), Night (18:00-24:00), representing by A, B, C, D respectively. According to result of ANOVA (Tab.6), there are significant differences in sections A-B, A-C, B-D, C-D (sig<0.05), showing that fluctuations in average deal price actually exist within 24-hour-cycle. In the early morning (A), deal price is low but starts to climb up from in daytime and reaches its peak at noon (B, C), and then descends in the evening (D). Similarly, the deal volume has also proved a significant difference within a day (A-C, A-D, B-D) in Tab.7.

We will further compress 336 hours data into a 24-hour-cycle diagram by calculating mean value of deal price in each hour, trying to be more precise about the law in one periodicity.

Tab.4 Average deal price Friedman test result

N	336
Chi-Square	336.000
df	1
Asymp. Sig.	0.000

a. Friedman Test

Tab.5 Test of homogeneity of variances

Levene Statistic	df1	df2	Sig.
4.838	3	332	0.003

Tab.6 ANOVA on average deal price per hour

TIME	PERIOD	Mean Diff	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
A	B	-5.95002 *	0.000	-8.391	-3.509
	C	-5.26136 *	0.000	-7.702	-2.821
	D	- .40265	0.746	-2.843	2.038
B	A	5.95002 *	0.000	3.509	8.391
	C	.68866	0.579	-1.752	3.129
	D	5.54737 *	0.000	3.107	7.988
C	A	5.26136 *	0.000	2.821	7.702
	B	- .68866	0.579	-3.129	1.752
	D	4.85871 *	0.000	2.418	7.299
D	A	.40265	0.746	-2.038	2.843
	B	-5.54737 *	0.000	-7.988	-3.107
	C	-4.85871 *	0.000	-7.299	-2.418

*. The mean difference is significant at the 0.05 level.

Tab.7 ANOVA on average deal volume per hour

TIME	PERIOD	Mean Difference	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
A	B	-10098.369	0.107	-22382.23	2185.49
	C	-19012.655*	0.003	-31296.51	-6728.79
	D	-26493.143*	0.000	-38777	-14209.28
B	A	10098.369	0.107	-2185.49	22382.23
	C	-8914.286	0.154	-21198.15	3369.57
	D	-16394.774*	0.009	-28678.63	-4110.91
C	A	19012.655*	0.003	6728.79	31296.51
	B	8914.286	0.154	-3369.57	21198.15
	D	-7480.488	0.232	-19764.35	4803.37
D	A	26493.143*	0.000	14209.28	38777
	B	16394.774*	0.009	4110.91	28678.63
	C	7480.488	0.232	-4803.37	19764.35

*. The mean difference is significant at the 0.05 level.

(3) Curve fitting

①Curve fitting for periodic average deal price:

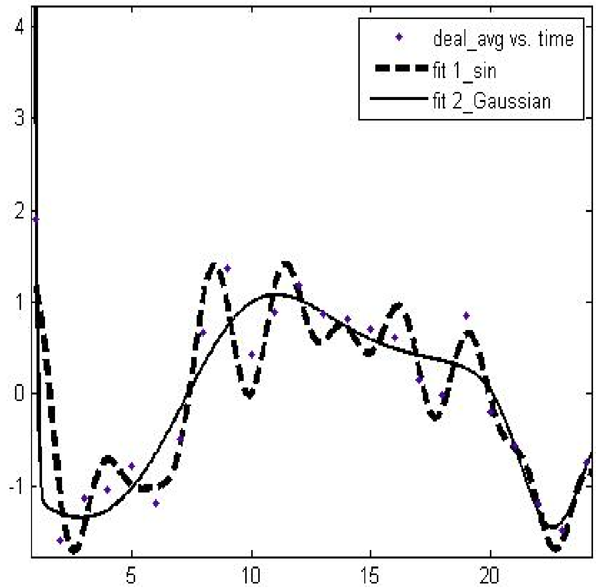

Fig.4 The average deal price curve fitting results

The consecutive line represents a Gaussian fitting, while the dashed line represents trigonometric fitting.

Gaussian mixture	Trigonometric
SSE: 2.185	SSE: 2.129
R^2: 0.905	R^2: 0.9074
Adjusted R^2: 0.7572	Adjusted R^2: 0.6452
RMSE: 0.4928	RMSE: 0.5957

It can be found that the average deal price starts to rise by 6:00, peaking at noon and began a significant decline in the evening.

②Curve fitting for periodic average deal volume:

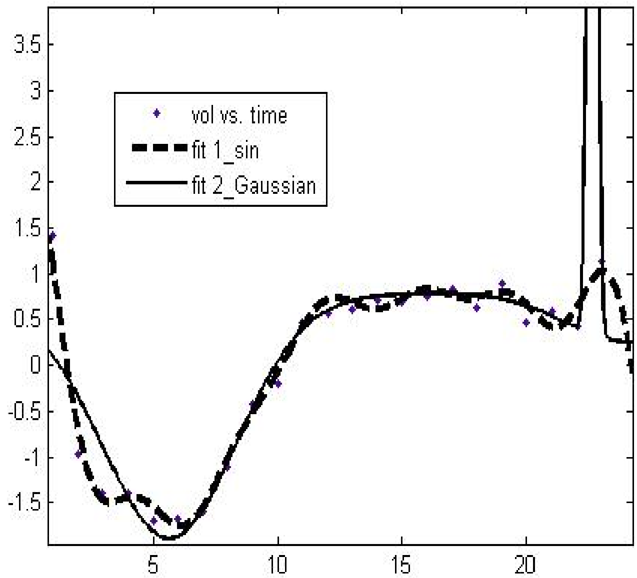

Fig.5 Average deal volume curve fitting results

The consecutive line represents a Gaussian fitting, while the dashed line represents trigonometric fitting.

Gaussian mixture	Trigonometric
SSE: 2.608	SSE: 0.4532
R^2:0.8866	R^2:0.9803
Adjusted R^2:0.7827	Adjusted R^2:0.9245
RMSE: 0.4662	RMSE: 0.2748

It can be found that the average volume curve is closer to the trigonometric curve, which means that the amount of impressions volatiles more regularly in one-day-time.

③ Curve fitting for periodic deal volume and price:

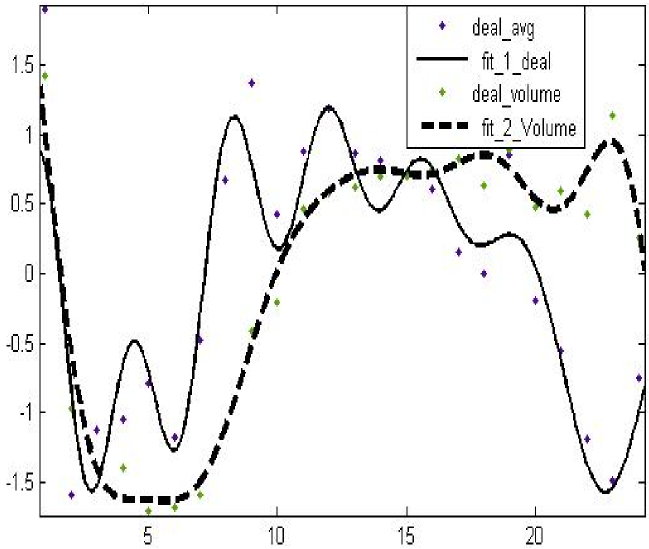

Fig.6 Periodic average deal price and volume curve fitting results

The consecutive line represents periodic deal price per hour, while dashed line represents period deal volume fitting.

It can be found that, in one day's cycle, deal price curve peaks at noon, prior to deal volume curve; while during intensive transaction period at around 20:00, prices are surprisingly declining.

(4) Summary of conclusions for average deal price, volume, click-through-rate per hour

We also put the average deal price, volume, CTR (Fig.7) into one periodic diagram in order to observe different performance in 24-hour time sections.

By analyzing data and images, we have reached the following conclusions:

(1)00:00 to 06:00 is characterized by **Low Price, Low impression, High CTR**

At this point, deal prices and impression volumes are in the low level, showing a "slack season" market; while click rate is not low, given that actual amount of ads displayed and the corresponding absolute times of clicks are small. This suggests that the early morning time is not entirely valueless, audience who stays up all night may also have potential effects.

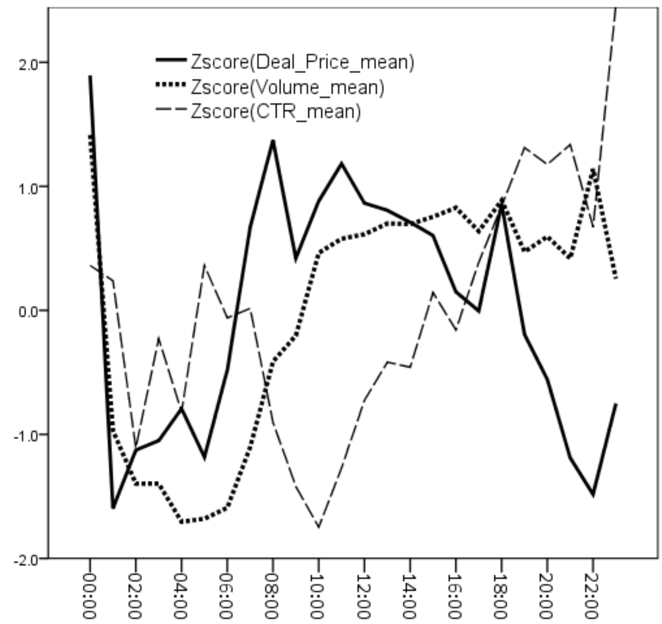

Fig.7 Periodic Average Deal Price, Volume & CTR curve

The consecutive line represents periodic average deal prices per hour in all the 14 days. The dotted line represents periodic average volume in each hour and the dashed line is on behalf of average click-through rate. The x-axis is a cycle of 24 hours; and all data has been Z-SCORE normalized.

(2)06:00 to 12:00 is **High Price, Low impression, Low CTR**

The average deal price and volume are both under trend of arising. The average deal price curve precede to reach a high area, while CTR continues to stay at low level. The number of bidding opportunities and impressions begin to climb after people get up, but the growth of clicks does not follow this trend, making click-through-rate decline. The average deal price is prior to reach its summit at noon, indicating that bidders are most competitive in this period, and actual effect of advertising may not be satisfactory.

(3)12:00 to 18:00 is **High Price, High Impression, low CTR**

Ad impressions continue to increase to a high-level in the afternoon and become stable; while the average deal price has started to decline; click-rate curve also begins to rise at this time, forming a trend of “restrain-then-rise”. This shows that although the average deal prices began to decline in the afternoon, but it still retains to be a competitive period. Due to gradually peaking ad impressions, CTR will be significantly more ideal and effect of advertising might be better.

(4)18:00 to 24:00 is **Low Price, High impression and CTR**

The average deal price further declines and both the volume and the click-through-rate reached their peak; especially in the traditional "prime time" (19:00 to 21:00), the average deal price is not at the high level of the day. It seems that in so-called prime time when ad impressions and CTR perform well, auctions on impressions are not competitive. In the traditional advertising period (i.e. the time during the day, from 8:00 am to 18:00), fierce competition may not produce ideal effects on the contrary. This shows that existing RTB advertising strategy still leaves space to be optimized.

In summary, we believe that in the period of A, B during 24-hour-period, average deal volume is low, but high volume in C, D periods; average deal price is lower in A, D and higher in B, C.

4 Bidding algorithm design

Based on the empirical analysis of RTB related datasets above, we propose a real-time bidding optimization algorithm in this section. RTB theory has proposed a standpoint of “audience buying”, instead of traditional “media buying”, or slot buying; therefore, deciding one’s bids through floor price set by publishers is inappropriate because it has no relationship with actual value of impressions in RTB circumstances. However, given the fact that information related to audiences or behaviors are limited when mining raw server logs, we cannot refer too much about audience. In that case, the macro-based auction performance is what we’re really concerned.

From the perspective of macro auction market, the characteristics of auctions are still associated with floor price and we are rightly utilizing this correlation, rather than floor price solely, to help design our real-time-bidding algorithm.

By further empirical analysis on floor price, and combining with previous observation on average deal price & volume, we presented a LEARN-THEN-BID algorithm.

4.1 Algorithms Goal

Initially, given the budgetary constraints, we intend to maximize the number of impressions n that we will win; the more impressions, the better. We define the number of impressions per unit budget i.e. budget utilization efficiency:

$$f = n / budget$$

Based on preferences or targeting group set by advertisers before auction, DSP will bid according to its client; thus it can be said that every impression to a potential viewer is valuable, because perhaps the user might produce clicks or conversions. Therefore, the bid reflects the value of this audience estimated by DSP. As an agency to advertiser, DSP should consume the budget as soon as possible and strive for each potential impression opportunity, rather than saving budget.

At the same time, the higher a bid is, the faster budget would be consumed. Sometimes, the high deal price may be raised up by intense competition, while sometimes due to reserved price of ad slots. For example, in large portals such as Yahoo or AOL, DSPs will compete fiercely on these ad slots (as for some clients, the times of impressions on these sites may still become an appraisal to advertising effects). Even though we won the bidding for this slot, the overall budget control is unfavorable and it is still media buying. That is why we propose the concept of budget utilization efficiency, trying to win as many impressions as possible, and hope to be more "efficient" in assigning budgets. This will help to evaluate the effects of algorithm from a broader point of auctions.

4.2 Reserve price and "price margin"

Most of the ad slots in RTB markets come from those small sites, i.e. "long tail" impressions. We've found that deals of impressions are to some extent associated with this trait. What we actually intend is thus to distinguish the two situations between long-tailed impressions and highly competitive ones. No doubt that reserved price α is one of the most convenient information we can utilize to differentiate long-tailed impression or not.

In order to verify this division, we will sort average deal price in every IMP dataset by descending and observe related data of floor price. Then we've noticed a phenomenon:

(1) For high reserved price cases, such as 295, 300 and so on, the gap, or margin, between deal and reserve price is not large, and it's quite often that the deal price equals to floor price;

(2) For low reserved price cases, the margin between deal and reserve price is larger and more volatile; for instance, the majority of long tail impressions with floor price of 0 or 5 may fluctuate greatly between 0-300.

We are very concerned about the price margin between deal and reserved price. Due to the oscillation, price margin can help us judge different strategies facing the two cases. For long-tailed impressions, the fact that price margin is more volatile can give us signal to enhance our bids in order to elevate the chances of winning them. By contrast, deal price for competitively bidding impressions cannot exceed greatly from its reversed price while competition would be even more horrible, which enlightens us that we may not have to make great efforts in bidding them and that we can also suitably reduce our biding prices in this case. After all, we should consider to allocate the given budget to more long-tail bidding opportunities, therefore winning more impressions and achieve our goal of budget utilization efficiency.

Although each incoming bid request would come up randomly, the margin between reserve and deal price can help us to develop bidding strategy facing every impression. If we can learn the average deal price based on historical datasets, then it is feasible to evaluate each of the incoming bid request arriving at different time points by adding a "margin" on the basis of average deal price previously learnt. At least, we can assure that each of our bid is larger than average deal price in the same time points.

4.3 Model assumptions

First, we made reference to some scholars **[9]** about **Fully Observable Exchange** assumption that the winner of the auction and the final deal prices are public on the market. Therefore we can directly study historical deal price levels rather than spending any budget.

Secondly, in our offline auction simulation, we have simplified the model that, apart from our own bid, the other party is max_price, representing the highest bid on the market among the rest of DSPs. We do not care who the highest bidder is but more concerned about: (1) whether we can win this bid, i.e. whether our bidding is higher than max_price; and (2) if wining the bid, how much the second price is. In other words, the market is reduced to a bipartite bidding market and we consider only the second price.

In addition, how to simulate max_price is an important step of the experiment. According to Tab.2, the average deal price per hour is around 75 (RMB/CPM) and the maximum is about 100. Therefore, we will set max_price as (1) a random number uniformly distributed from 0 to 100, (2) fixed sum of 300 (RMB/CPM), and (3) a random number uniformly distributed from 0-400 (RMB/CPM).

Finally, as for the budget, we consider the condition that budget is adequate. As a bipartite model, the sum of initial budgets for two players should be entire transaction values in the market. Although we initially consider adequate budget situations, due to the randomly generated deal price, budget overruns can hardly be avoided. So we require that: if spends from both sides on the market is less than budget, i.e. expend <budget, the auctions can continue as usual. However, if either party runs out of budget, i.e. expend> budget, then the player who overruns can no longer bid, and all remaining impression opportunities will be assigned to the rest party until the bidding process is finished. We refer to this rule as "left-take-all". There may exist an extreme case that a party is always winning and all of the impressions will be assigned to him until all his budget is consumed. At this time, if the bidding process is stopped manually, the other party who always fails will exit without spending a ratio of spending and there still remains impressions. To avoid this situation, we therefore developed "left-take-all" rule, which is fair to both sides.

4.4 Learn-then-bid

Based on IMP log from a certain day in March, we can learn historical average deal prices and volumes on open market. For each hour t, we define the average deal price as avg_deal (t), whose characteristic is in line with our empirical results that low average deal price in the morning and night and relatively higher in daytime. The avg_deal (t) can be referred as BasePrice.

We experiment on potential BID datasets from a different day in June as source of RTB auctions. Note the arrival of each impression as i. It is known that the required floor price for i is α(i), and the bidding comes at hour t. Due to historical average deal price during same hours in the past, we can calculate a margin price between deal and reserve price:

$$d = avg\_deal(t) - \alpha(i)$$

Note that d can be determined during each impression i.

If d> 0, it means that we will bid on top of the BasePrice by adding this margin. At the same time of t, d will increase when meeting a long-tail impression since reserved price is lower; thus the bidding price is enlarged, elevating the chance to win corresponding impression without intensive competition. On the contrary, as for higher reserved price case, margin price d becomes smaller and at the same time of t, bidding price will add only a small value. The main reason to do so is to avoid the high competition by bidding a relatively low price for non-long-tailed impressions and to allocate the identical budgets to win additional impressions with less competition, which will achieve our experiment goal f, i.e. an optimized utilization of funds.

If d <0, we do not bid in the current experiment. Those impressions exceeding average deal price does not belong to long-tail resources we are caring about. They account for only around 1% among all of the 14 IMP datasets. Therefore, we believe that temporarily ignoring these bids has no prejudice to the outcome of the auctions, and that it is also in line with our fight to win the long-tail impressions.

But is d = avg_deal (t) - α(i) a reasonable choice? Is 2d, 3d, etc. more optimal? And we don't know whether or not 0.5*d may obtain the optimized utilization of budget f. In that case, an additional control parameter λ is determined to judge the effect of different margins on the final number of impressions. We finally define our bid price as:

$$\text{Bidding Price} = avg\_deal(t) + \lambda d$$

Theoretically, a bid on the basis of BasePrice, plus a margin d related with reserved price, will achieve: (1) during competitive bidding period t, i.e. from 06:00 to 18:00, the BasePrice itself is high. When encountering long-tailed impressions, the margin d becomes larger so that bidding price will be further enhanced, which is also beneficial to win more impressions in these highly competitive hours and simultaneously avoiding great consumptions on the budget. (2) In less competitive hours t, i.e. from 00:00 to 06:00 and 18:00 to 24:00, the BasePrice is not high but our overall bids are still above historical average deal price. By avoiding high competition, not only can we win additional times of long-tailed impressions, the budget can also be allocated to hours in the afternoon or evening when deal volumes are greater and CTR is more ideal.

4.5 Experimental results and analysis

Assume the initial budget budget equals to 100000000, representing the both sides enjoy adequate budgets. The total times of experiments or in-coming i are 1,736,876. We take λ = 1,2,3,4, separately, and then run offline bidding programs. The results are recorded in the table below (Tab.8). We note n as the number of impressions we have won and n_for_DSPs as representative of impressions other bidders have won in the market.

Tab.8 Experimental results

	Max_price	N	n_for_DSPs
λ=1	Uniform(0,100)	1,709,940	26,936
	300	1,091,593	645,283
	Uniform(0,400)	1,076,169	660,707
λ=2	Uniform(0,100)	**1,710,672**	26,204
	300	**1,304,603**	432,273
	Uniform(0,400)	**1,270,211**	466,665
λ=3	300	333,334	1,403,542
	Uniform(0,400)	643,702	1,093,174
λ=4	300	333,334	1,403,542
	Uniform(0,400)	537,474	1,199,402

When max_price, is a random value uniformly distributed from 0 to 100, our algorithm is overwhelming, because by adding a "margin price" to BasePrice, we can always offer a bid which is generally greater than 100 but less than 300. The max_price side will not only lose, but also cause the second price to be relatively small, which cannot bring much pressure on my budget and I can easily absorb all of the impressions within a low payment. Increasing the value of λ does not change this situation, and therefore it's useless to continue with this max_price. Note that the average deal price per hour is only around 75 (RMB/CPM), as previously mentioned.

We continue the simulation to assume max_price as a fixed value of 300. This bidding comes from the DSP who provides data. When λ = 1 and λ = 2, the opponent wins all impressions from beginning; but, due to my high bids, he quickly consumed his budget. And the rest of impressions will become my property. While λ is larger, my bids will be higher and the greater the opponent needs to pay. The cost will be maximized if λ = 2. However, we can state that higher bids are not always appropriate. When λ = 3 and λ = 4 or higher, the situation has flipped that I've succeeded in almost the entire bidding process from beginning until I consumed budgets prior to my rival. In fact, due to the second price becomes a fixed 300, no matter if the bidding price rises, I can only win up to 333,334 (=100,000,000/300) times of impressions. This proves that overestimated bids won't contribute but reduce budget utilization efficiency f.

Finally we set max_price to be uniformly distributed from 0-400. Then an input of $\lambda = 2$ would still achieve maximum impressions, even rising my bids. The number of impressions won will instead decrease, thus jeopardizing f. On the other hand, if taking λ less than 1, not only will we lose each auction, but the rival is under minor pressure of payment. At this point, we can believe that the budget utilization efficiency f will converge when $\lambda = 2$, which is an optimal solution.

Through our algorithm and experiments conducted offline, we believe that it is a feasible strategy to achieve the goal of budget utilization efficiency f by bidding on the basis of historical BasePrice from marketplace in different time points (i.e. avg_deal[t]), plus a dynamic margin price accustomed to every auction case. This method not only ensures that almost each bidding price is higher than overall avg_deal[t], which can be learned in past hours, but also that price margin d can be intelligently adjusted, accordingly with the general phenomenon concerning deal and floor price. Second, an overly offered bid is not economical. An overly-estimated bid may indeed raise the cost of payment for the rival, but we can also be jeopardized by the cut-throat competition, losing some of the long-tailed impressions and thus reducing efficiency f.

We suggest that bidders should be loyal to their original evaluation of the impression and bid reasonably to avoid fierce competition, which is also essential and beneficial to the general RTB market. Especially when considering the empirical results we've found, it is crucial to protect the efficiency of expends since both the average deal volume and CTR per hour peaks at the time of afternoon and evening.

5 Conclusions

By analyzing raw server logs recording bidding, impression and clicks from a domestic DSP on fourteen days, this paper discusses the relationship between average deal price, volume and click-through-rate in different time periods. Our conclusion is:

(1)00:00 to 06:00: the dawn section presents low deal price, low deal volume but high CTR;

(2)06:00 to 12:00: the morning section shows high deal prices, low volume & low CTR;

(3)12:00 to 18:00, the afternoon session presents high deal price, high volume but low CTR;

(4)18:00 to 24:00, the evening session shows low deal price, high volume & high CTR;

(5)Depending on the time period, bidders can make different strategies from traditional advertising of "buying media"; bidding offers can be adjusted according to the characteristics of deals and impressions in a smart way while an excessive bid will not help DSP gain more impressions.

There still exist some other works to do in the future. First, the data is limited with only 24 * 14 hours, we need more continuous data to further validate the reliability of our empirical analysis. In addition, our algorithm experiments simply abstract the highest bid from other participants on the market by uniformly distributed values. In fact, based previous studies by other scholars **[9]**, the winner bids from current Ad Exchange (such as Right Media Exchange) are subject to the lognormal distribution; we can also add this part of simulation in subsequent experiments. Finally, there remains the problem that, although those long-tailed impressions may account for the majority, why is it that fierce competition is taking place during daytime? Further explanation may also be required for the fact that prime period at night has low level of deal price while the deal volume and CTR are performing well.

This article only empirically analyzes the patterns embodied inside datasets and further realizes a simplified and preliminary bidding algorithm. Later, we will consider the utilization of time series analysis methods, such as ARMA, GARCH models to provide more specialized data analysis on deals and auctions, and even expecting to refer many theoretical studies on stock exchange markets. We sincerely hope that, with concept that "advertisement is commodity", our finding concerning the overall and macro-based real-time-bidding patterns or conclusions can provide some new ideas on the research of Real-Time-Bidding algorithms.

References

[1]IDC-Research: Understanding real-time bidding (RTB) from the publisher perspective.

[2]IDC-Research: Worldwide and US real-time bidding 2012–2016 forecast– IDC Q32012.

[3]Goldstein D G, McAfee R P, Suri S. Improving the effectiveness of time-based display advertising [C]// Proceedings of the 13th ACM Conference on Electronic Commerce. ACM, 2012: 639-654.

[4]Mansour Y, Muthukrishnan S, Nisan N. Doubleclick ad exchange auction [J]. Arxiv Preprint ArXiv, 2012: 1204.

[5]Muthukrishnan S. Ad exchanges: Research issues [M]. Internet and Network Economics. Springer Berlin Heidelberg, 2009: 1-12.

[6]Muthukrishnan S. AdX: A model for ad exchanges [J]. ACM SIGecom Exchanges, 2009, 8(2): 9.

[7]Vickrey W. Counterspeculation, auctions, and competitive sealed tenders [J]. The Journal of Finance, 1961, 16(1): 8-37.

[8]Chen Y, Berkhin P, Anderson B, et al. Real-time bidding algorithms for performance-based display ad allocation [C]// Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011: 1307-1315.

[9]Ghosh A, Rubinstein B I, Vassilvitskii S, Zinkevich M. Adaptive bidding for display advertising. In Proceedings of the 18th International Conference on World Wide Web, 2009: 251-260.

[10]Perlich C, Dalessandro B, Hook R, et al. Bid optimizing and inventory scoring in targeted online advertising [C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012: 804-812.
[11]Panaligan Reggie A C, Chen A. Quantifying movie magic with google search. Google Whitepaper Industry Perspectives User Insights, 2013.
[12]Li W, Wei Q, Chen Y. Regional analysis on chinese internet ads with click-through rates [C]// Proceedings of the 1st International Conference on E-Business Intelligence (ICEBI2010) Atlantis Press, 2010.
[13]Chatwin R E. An overview of computational challenges in online advertising [C]// American Control Conference (ACC), 2013. IEEE, 2013: 5990-6007.
[14]Lu P, Hua Teng S, Yu C. Truthful auctions with optimal profit. Workshop on Internet and Network Economics (WINE 2006), 2006: 27-36.
[15]Mirrokni V, Muthukrishnan S, Nadav U. Quasi-proportional mechanisms: Prior-free revenue maximization. Arxivs, 2009.
[16]Myerson R B. Optimal auction design. Mathematics of Operations Research, 1981, 6(1): 58-73.
[17]Lahaie S, Parkes D, Pennock D. An expressive auction design for online display advertising. AAAI, 2008: 108-113.
[18]Chakraborty T, Even-Dar E, Guha S, Mansour Y, Muthukrishnan S. Callout optimization, 2009.
[19]Ghosh A, McAfee P, Papineni K, Vassilvitskii S. Bidding for representative allocations for display advertising. Leonardi, S. (ed.) WINE, 2009, 5929: 208-219.
[20]Klemperer P. Auction theory: A guide to the literature. J. of Economic Surveys, 1999: 227-286.
[21]Madhavan A. Market microstructure: A survey. Journal of Financial Markets, 2000, 3(3): 205-258.