Individual report

Zhilin Cai
zhilin.cai.16@ucl.ac.uk

1. INTRODUCTION

Real-time bidding (RTB) has enjoyed a great popularity in online advertising these years due to its efficiency. Buyers bid in the real time to buy their desired ad slots in RTB. Since this only happen in a very short time, advertisers need a bidding strategy to find the most valuable impression for them and bid a proper price. Bid is automatically made in the demand-side platform (DSP) [3].

In this paper, an impression data set is investigated in detail in order for evaluation of basic bidding strategies including constant bidding, random bidding and linear bidding, which will be included in the group report of my team. Then, I will give a comparison of basic bidding strategies and optimal real-time bidding (ORTB) strategy. Finally, an optimal linear bidding strategy will be proposed, which is considered as the best bidding strategy for our group. Related code is uploaded to Github: `https://github.com/edcys/COMPGW02-Web-Economics`

2. RELATED WORK

An advantage of RTB is that it could allow advertisers to provide a bid for each individual impression, which is distinct from the conventional negotiation or presetting a fixed bid [1]. In data exploration of this report, I will refer to Zhang's paper [3]. The idea and the parameters that we analyze and evaluate the benchmark would also refer to this paper [3], which illustrates the way to estimate estimated CTR and analyses the result of evaluation. While for the part of best bidding strategy, I will evaluate ORTB in this paper, which is considered as the most effective bidding strategy in RTB [2].

3. DATA EXPLORATION

3.1 Basic Statistics

Generally, training set and validation set relatively include data of 2697738 and 299749 ad impressions from 9 advertisers including 26 features. Since different advertisers are from different industry according to Zhang [3], it can be meaningful to separate data correspond to each advertiser in the investigation. Table 1 and 2 give detailed statistics of training data and validation data about the main metrics used in evaluation of bidding strategies, including Click-Through Rate (CTR), Clicks (Number of Clicked Bids), Cost (Total Money Paid), CPM (Cost Per Mille) and eCPC (Average Effective Cost Per Click).

Table 1: Training data statistics

Adv.	Clicks	Imps	Cost	CPM	CTR	eCPC
1458	451	540293	37231	68.91	0.083%	82.55
2259	45	146778	13649	92.99	0.031%	303.31
2261	37	120619	10789	89.45	0.031%	291.60
2821	144	231416	20625	89.13	0.062%	143.23
2997	251	54487	3413	62.64	0.461%	13.60
3358	233	304782	28145	92.35	0.076%	120.80
3386	358	498554	38341	76.90	0.072%	107.10
3427	340	454031	36820	81.10	0.075%	108.29
3476	175	346778	27481	79.25	0.05%	157.04
Total	2034	2697738	216496	80.25	0.075%	106.44

Table 2: Validation data statistics

Adv.	Clicks	Imps	Cost	CPM	CTR	eCPC
1458	50	60025	4139	68.96	0.083%	82.78
2259	11	16419	1519	92.55	0.067%	138.15
2261	5	13370	1196	89.47	0.037%	239.25
2821	16	25632	2281	89.01	0.062%	142.59
2997	26	6034	387	64.20	0.431%	14.90
3358	27	33853	3125	92.34	0.08%	115.77
3386	33	55196	4255	77.10	0.06%	128.95
3427	45	50381	4077	80.93	0.089%	90.61
3476	13	38839	3062	78.85	0.033%	235.58
Total	226	299749	24045	80.22	0.075%	106.39

As can be seen, advertiser 2997 enjoys the highest CTR among all advertisers and much more than the average CTR. The reason is that advertiser 2997 focus on mobile e-commerce app install ads, which is more likely to be clicked by accident according to "Fat finger" effect. Although the average price that paid on an impression for all advertisers is similar, indicated by similar CPM for all advertisers, their eCPC vary from each other. It means ad markets in different industries are different, which indicates the necessity of separating the data by advertisers.

3.2 User Feedback

Figure 1 gives an illustration of CTR distribution against six features for advertiser 1458 and 3358 in training data. In general, different values of features can make great difference on the performance of CTR.
Weekday: advertiser 3358 gets a very low CTR on Tuesday and it increases to the peak when it comes to Thursday. While CTR of advertiser 1458 seems stable relatively.
Hour: CTR for advertiser 1358 and 3358 keep low except the case that CTR of advertiser 3358 increases significantly at night.

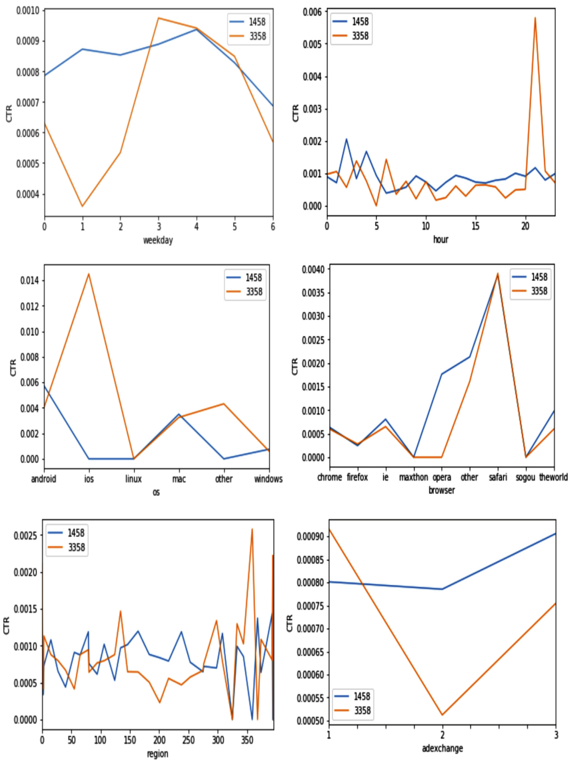

Figure 1: CTR distribution against different features for advertiser 1458 and 3358.

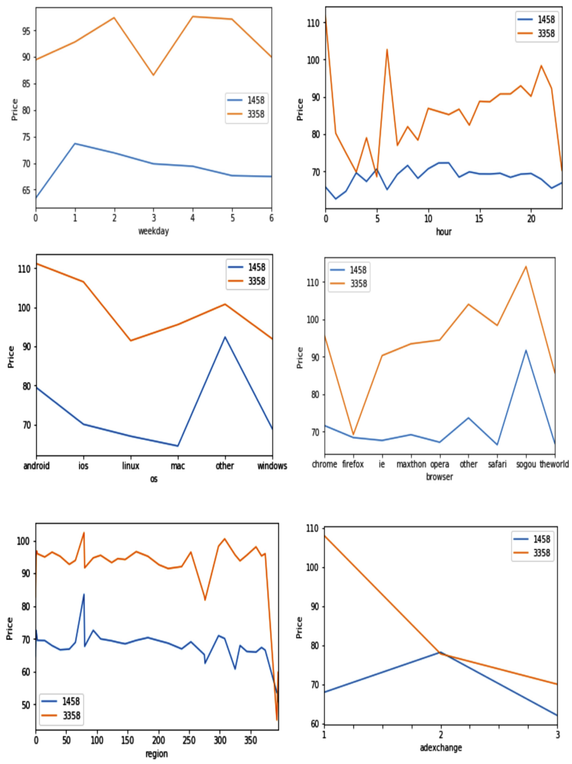

Figure 2: Market price distribution against different features for advertiser 1458 and 3358.

OS: Ad from advertiser 3358 enjoys high CTR on mobile device while advertiser 1458 does not. The possible reason can be seen from their industry category: advertiser 3358 is about software while advertiser 1458 is about Chinese vertical e-commerce.
Ad exchange: CTR varies on exchanges because different exchanges are connected by different publishers according to Zhang [3].

In summary, the analysis above indicate that to predict CTR for an impression, models for each advertiser need to be trained accordingly.

3.3 Bidding Behaviour

As second price auction is employed in the data set, a bid with highest bid price win the auction and pay the second highest pay price, namely the market price. A higher pay price indicate a more competitive environment of the market. Market price distribution against six features for advertiser 1458 and 3358 is depicted in Figure 2. In general, advertiser 3358 bids higher price compared to advertiser 1458 on average.

3.4 eCPC

There can be cases with a low CTR but a high cost when the two figures above were observed together. For instance, advertiser 3358 has a poor CTR on Tuesday but the cost is still high, which means it is not cost-effective. This is be measured by eCPC as shown in Figure 3. Figure 3 depicts eCPC against six features for advertiser 3358. As can be seen, click on ads from advertiser 3358 displayed on windows is the most cost effective compared to other operation systems. Advertisers can focus on the features that can bring them high eCPC and evaluate their bidding strategies, such as allocate more budget to bid on those ad positions. Analysis above indicates that different features can influence user feedback and market price so it is necessary to train models to predict CTR for each advertiser accordingly.

4. BEST BIDDING STRATEGY

4.1 Optimal real-time bidding strategy (ORTB)

To develop a better bidding strategy, I firstly try to implement the Optimal Real-time Bidding (ORTB) strategy proposed in Zhang's paper [2]. The fomula used is

$$b_{ORTB}(\theta) = \sqrt{\frac{c}{\lambda}\theta + c^2} - c \tag{1}$$

,where θ is pCTR and c and λ are the parameters to be optimized. According to Zhang, I try c in [20, 50, 80] and λ in $[1^{-7}, 5^{-7}, 1^{-6}, 5^{-6}, 1^{-5}, 5^{-5}]$ and tune parameters for each advertisers independently. However, I find out that ORTB performs even worse than linear strategy. There might be two possible reasons: Firstly, it is a special case under 1/4 budget limit. As can be seen from Figure 4, ORTB outperforms Lin within 1/2 and 1/8 budget limit but performs worse than Lin in 1/4 budget. The second reason might be because that when developing Lin and ORTB , I distributed budget to each advertiser according to their cost portion, which is talked in detail in the group report. While it might not be a optimal method since it does not consider more

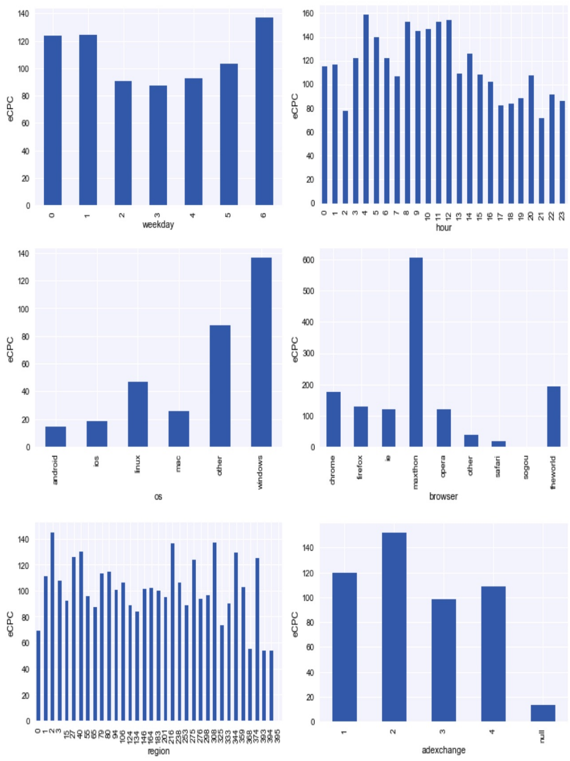

Figure 3: eCPC against different features for advertiser 3358.

important metrics such as Clicks, CTR or eCPC when distributing the budget. Inspired by that idea, I try to change the way of distributing budget by eCPC in training data: if an advertiser have a lower eCPC, it will get more budget. The specific budget for each advertiser is calculated by the following formula:

$$budget\_of\_Adv = total\_budget \times \frac{eCPC\_of\_Adv^{-1}}{\sum eCPC\_of\_each\_Adv^{-1}} \quad (2)$$

The result of the optimal ORTB is shown in Table 3. As expected, new distribution method optimized ORTB.

4.2 Optimal linear bidding strategy (OLB)

Table 3: ORTB strategy (budget distributed by eCPC) (1/4 budget)

Adv.	Clicks	Imps	Cost	CPM	CTR	eCPC
1458	38	41300	2805	67.94	0.092%	73.84
2259	3	5751	236	41.09	0.052%	78.78
2261	3	6101	309	50.75	0.049%	103.22
2821	2	5596	155	27.77	0.036%	77.70
2997	16	4143	263	63.64	0.386%	16.48
3358	6	6187	237	38.45	0.097%	39.65
3386	7	10668	316	29.69	0.066%	45.25
3427	22	22707	1211	53.38	0.097%	55.09
3476	10	15008	712	47.47	0.067%	71.24
Total	107	117461	6249	53.21	0.091%	58.41

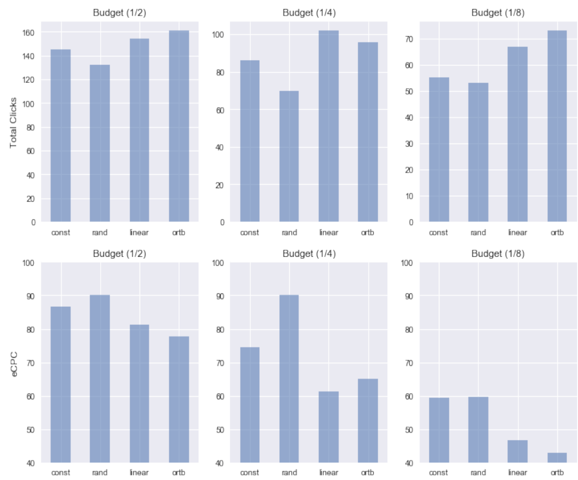

Figure 4: Strategies performance under different budget limit

Table 4: Optimal linear bidding strategy performance (1/4 budget)

Adv.	Clicks	Imps	Cost	CPM	CTR	eCPC
1458	35	48846	2594	53.11	0.072%	74.12
2259	4	4380	135	31.01	0.091%	33.95
2261	3	6416	258	40.31	0.047%	86.20
2821	5	8150	259	31.84	0.061%	51.91
2997	22	5425	284	52.37	0.406%	12.91
3358	8	8095	334	41.32	0.099%	41.81
3386	10	10569	335	31.76	0.095%	33.57
3427	23	21602	1027	47.58	0.106%	44.69
3476	10	15940	740	46.47	0.063%	74.07
Total	120	129423	5970	46.14	0.093%	49.76

Despite of the improvement on ORTB, ORTB_eCPC just slightly outperforms Lin_cost in 1/4 budget constraint. Therefore, I try to apply the new budget distribution method on linear bidding strategy. Surprisingly, the result of new linear bidding strategy nearly dominates ORTB except advertiser 1458. From the joint observation on Table 3 and 4, optimal linear bidding strategy (OLB) has more Clicks, lower Cost, lower CPM, higher CTR and lower eCPC in total and for most advertisers, except that a slight worse performance for advertiser 1458.

5. CONCLUSION

In this report, I made a data exploration on the ad impression data set, especially in the aspects of basic statistics, user feedback, bidding behaviour and eCPC. Then, an optimal linear bidding strategy (OLB) within 1/4 budget constraint was proposed. In my experiments, OLB outperforms other bidding strategies such as ORTB. However, I did not experiment the comparison under other budget constraint finally. In the future, I can try to investigate the influence of budget limit on the evaluation of different bidding strategies.

In the process of doing this project, Yuncan Zhang, Yao Fu and I worked together to finish the code and group report. My job was to evaluate each bidding strategies and share a performance report. I also did the related write-up about introduction and Problem 3: Linear Bidding Strategy in the group report. Yuncan Zhang wrote Problem 2: Basic Bidding Strategies and Yao Fu wrote Problem 4: Your Best Bidding Strategy in the group report. The rest of the group report is completed by us together.

6. REFERENCES

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

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

[3] W. Zhang, S. Yuan, J. Wang, and X. Shen. Real-time bidding benchmarking with ipinyou dataset. *arXiv preprint arXiv:1407.7073*, 2014.