A SURVEY ON BIDDING OPTIMIZATION TECHNIQUES IN ONLINE ADVERTISING

[1]Anjesh Kumar, [2]Sukrant Kathuria, [3]Shikha Gupta, [4]Atul Mishra
[1,2]Research Scholar, [3]Assistant Professor, [4]Professor

Abstract : Online Advertising is the main source of income in Digital Market as the internet is used widely today. Although, there are so many ways for advertising but because of the popularity of the internet, digital advertising became very popular and effective to earn revenue. But it also faces many problems. Bidding optimization is one of the most crucial problems in online advertising. Keyword-level bidding techniques are used in Sponsored search (SS) auction processes, because of the randomness of user's query behavior and platform nature. But the display advertising (DA) has taken benefits of real-time bidding (RTB) process to improve the performance for advertisers. In our work, we consider the RTB problem in sponsored search (SS) auction, named as SS-RTB. SS-RTB has a more complicated dynamic behavior, due to stochasticness of user's query behavior and more complicated bidding prices policies based on multiple keywords of an advertisement. Most previous techniques for Digital Advertisement can't be applied. Here analyze a reinforcement learning (RL) method for handling the complex dynamic behavior of bidding. Here, the state transitions fluctuate between two days. By observation of sequences, formulate a robust MDP model according to time i.e., hour-aggregation level of the auction's data and developed a model for SS-RTB. By not generating bid prices directly, Robust MDP model is used for impressions of hours basis and performed real-time bidding.

IndexTerms **- Real-time bidding, computational advertising, Reinforcement Learning (RL), Sponsored Search (SS), Markov Decision Process (MDP).**

I. INTRODUCTION

Real time bidding (RTB) technique is a rising and encouraging business model of online digital advertising markets. It also represents the main model of the computational digital advertising research, and the third world widely used advertising model following the display advertisement network and keyword search-based advertising. Via online generated cookies-based analysis of big data, RTB has developed potential of identifying the quality and interest of the audience behind each ad conversions, and then delivers the best keyword basis matched ads accordingly.

As such, RTB is the main reason of a transformative innovation in online digital advertising markets. It evolved from the traditional advertisement method of "media buying" and "ad slot buying" patterns to the unstructured big data driven "audience buying" pattern. In other words, RTB evolved from large scale wholesaling into customized or selective retail when selling ad impressions online, which significantly increases the accuracy and efficiency of various ad delivery technologies.

Real time bidding (RTB) advertising has experienced a heavy growth. In the international digital markets, it is reported that 88 % of North American advertisers have started to use RTB when buying ad impressions in 2011. In China, the RTB market starts from the TANX system in 2011. According to report analysis in China in 2013, the RTB amount ad delivery has crossed 5 billion impressions and advertiser's RTB budgets was increased by 300 % to 83 million dollars [1].

The real-time bidding (RTB), also known as programmatic digitally buying, has recently become the fastest rising area in online advertising. Instead of bulking buying, RTB utilized various computer algorithms and techniques for automatically buying and selling their ads in real time scenario; RTB uses per impression context and targets the different ads to only specific people based on real time data like geographic location, history of previous searches, and hence dramatically increases the productivity of display advertising. Here, business process of RTB ad Delivery and key roles in RTM markets are explained. Using the data sampled from both demand and supply side, an emerging selling infrastructure can be developed and it can also help to identify various research and design defects or issues in such real time systems. It is also observed that periodic patterns occur including impressions rates, clicks rates, bids rates, and conversion rates which also suggested that time dependent models can be appropriate for monitoring the repeated patterns in RTB scenario. It also found that despite the claimed 2nd price auction, the 1st price auction payment in fact is accounted for 55.4% of total cost due to the designing of the soft floor price [2]. As such, we can argue that the setting of floor price i.e., soft and hard floor in the current RTB systems puts publisher in a more favorable position and advertisers in a less favorable position. Furthermore, different analysis on the conversation rates explained that the current bidding techniques used are far less optimal, thus, required the deep changes for optimization in algorithms considering the facts such as the temporal behaviors, the recency and frequency of the various ad displays, which have not been well considered in the previous time.

Bidding optimization algorithms is required for maximizing the advertiser's profit or revenue in online advertising. In the sponsored search (SS) scenario, optimization problem is typically formulated advertisers' objectives via searching the best settings of keyword level bids [10, 13]. The keyword level bids are usually considered to be static during the online auction process. However, the pattern of user queries develops a complicated dynamic environment where a bidding strategy could significantly increase advertiser's revenue. It became more important on e-commerce websites auction platforms since

impressions are converted into purchases. Thus, Sponsored Search Real-Time Bidding (SS-RTB) aims to generate effective bids at the impression level in the context of Sponsored Search.

II. LITERATURE SURVEY

Online advertising has been widely studied in recent years, where many researchers have developed solutions to the various problems from all the auctions perspective.

Edelman et al. [3] studied the various properties of the generalized 2nd price auctions, which has been extensively utilized by many search engines to sell online ads. In GSP (generalized second price) auction, for a specific set of keywords/keyword, advertisers submit their bids stating their maximum willingness to pay for a single click. When a user search for a keyword, he/she receives search results along with sponsored links, the latter shown in decreasing order of bids.

Aggarwal et al. [4] proposed a truthful auction for ad slot advertising on webpages. It presented a truthful auction for pricing in real time bidding on a web page assuming that ads for different merchants ranked in decreasing order of their weighted bids. This capture both the Overture model where bidders are ranked in order of the submitted bids, and the Google model where bidders are ranked in order of the expected revenue (or utility) that their advertisement generates. Assuming separable click-through rates, we prove revenue-equivalence between auction and the non-truthful next-price auctions currently in use.

Cary et al. [5] studied advertisers' bidding strategies for a repeated keyword auction, by considering several properties including revenue, convergence and robustness.

Arnosti et al. [6] studied how to allocate impressions efficiently in online advertising when advertisers have correlated valuations and are differentiated in abilities to estimate the values of impressions. Besides ad auctions, allocating ads through contracts has also been studied by a few researchers.

Hu [7] studied the profitability of performance-based pricing models in online advertising, and proposed an optimal contract to maximize the total utilities. it focuses on two entities that are involved in an online advertising contract: an online advertiser and an online content publisher. The advertiser sells a product (or service) to consumers through the online channel. In order to boost its sales, the advertiser launches an online advertising campaign by designing an advertisement and contracting with a publisher so that the publisher would deliver its advertisement to consumers who may be interested in the product it sells. An impression is meant to be instance of the ads being served to a customer's search page results. Every time the advertisement is served to a consumer's browser, the consumer may choose to ignore the advertisement, or to click on the advertisement which can be refer as a click through. If the consumer is taken to the advertiser's online store, the consumer may make a purchase, or leave without making a purchase. Click-through rate (θ_c) is defined as the ratio of click through to impressions, and purchase rate (θ_p) is defined as the ratio of purchases to impressions.

Moon and Kwon [8] considered a contract problem between advertisers and publishers, and proposed a hybrid pricing scheme based on both cost per impression (CPM) and cost per click (CPC). However, none of these works jointly considered the ad auction and the contract, and thus their solutions cannot be applied to scenarios where the ad allocation involves both of them.
Among various ways of pricing schemes online advertisements, the methods is based on cost-per -impression (CPM) and cost-per-click (CPC) are the two most popular. The CPC fee is directly proportional to the click through rate (CTR)i.e., the ratio of the number of times clicks on ad to the number of times it shown to the users, which is uncertain and makes decisions of advertisers and publishers difficult. So used a hybrid pricing scheme i.e., advertisers pay the minimum of cost per impression (CPM) and cost per click (CPC) fees by purchasing an option from publishers. Nash bargaining game for negotiation between an advertiser and a publisher to decide option price and provide the solution. Furthermore, it shows that such option contracts will help the advertiser avoid high cost for ads and the publisher generate more profit. Compared to CPC and CPM, the option contract will improve the feasibility of the model.

Amin et al. constructed [9] used a Markov Decision Process (MDP) for budget optimization in SS [11]. This MDP method deals with impressions and different auctions held in a batch model and hence can't be used for RTB Environment. Moreover, the underlying environment for MDP is "static" in that all states share the same set of transition and they do not consider impression specific features. Such an MDP cannot well capture the complex dynamics of auction sequences in SS, which is important for SS-RTB.

Cai et al. developed a reinforcement learning (RL) method for RTB in DA [12]. The method combines an optimized reward for the current impression (based on impression-level features) and the estimate of future rewards by an MDP for guiding the bidding process.

III. BUSINESS MODEL OF THE RTB MARKETS

Here, we briefly explain the key roles of various parameters in RTB markets and their typical business process of RTB advertisement delivery.

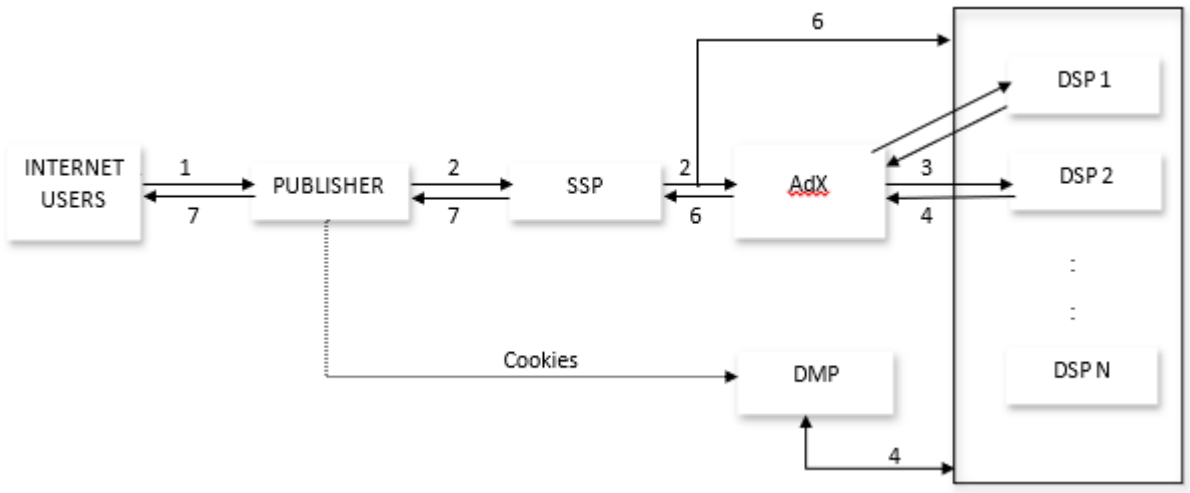

Figure 1. RTB Business Process Ad Delivery

3.1. The Key Roles in RTB Markets

The key roles of various parameters of the RTB market includes the publisher, advertiser, SSP, AdX, DMP and DSP. Each role is explained as follows:

- **Publisher** is the website's owner. When a user visits a publisher's website will activate an impression and the winning advertiser's ads of the RTB auction process on this impression will be exhibit on the publisher's webpages.
- **Advertiser** is the buyer of advertisement impressions. In RTB auctions process, advertisers bid for ad impressions on keyword basis according to their marketing objectives, budgets constraints, strategies and policies etc. The advertiser with the maximum bid wins the ad impression.
- **SSP** is also known as supply side platform. It is an agency platform which aims to help publishers to optimize the strategies and policies of managing and pricing their advertisement inventory, including setting of the optimal reserve prices, allocating ad impressions to various channels, etc.
- **AdX** is an ad exchange market that is used to match the sellers and buyers for each impression. AdX uses standard protocols to provide the various ad requests and user information among other key elements in RTB markets, which aim to find the best keyword basis match of advertisers and their target audiences. Thus, it plays a most important role in RTB markets.
- **DMP** is also known as data management platform that collects, stores and analyzes the cookies data generated by the browser of Internet users. It provides DSPs and AdXs paid services of identifying and targeting their audiences
- **DSP** also known as demand side platform and is an agency platform that provides advertisers to optimize their strategies, policies of ad management and their ad delivery. Based on its big data analysis in recent years and audience targeting technologies based on many factors, as well as its RTB environment and algorithms, DSP helps advertisers buy the keyword level best matched ad impressions from various AdXs in an easy, convenient and unified way.

3.2. RTB Business Process Ad Delivery

As can be seen in Figure 1, the typical RTB business process ad delivery can be depicted as follows [l].

1. An Internet user visits the publisher 's webpages.

2. There are many ad slots for selling in the RTB markets, an ad request is sent by publisher to the AdX and SSP, which contains the information of the user, ad slots with their reserve prices.

3. AdX will forward the information to all eligible DSPs after receiving the ad request from SSP.

4. Each DSP then parses this information and asks the DMP about the required information of this user (e.g., its geographic location, age, gender, historical behavior, interests of shopping and intentions, etc.), and starts an auction accordingly to match advertisers. The winner advertiser's ad will be fed back to the AdX, and each DSP response within a specific preset time.

5. AdX starts an auction to find winner advertisers from each DSPs, and determines the maximum bid among these winner advertisers. If this bid is lower than the reserve price of publisher, this RTB process will terminated the ad slots empty or reassigned to non RTB channels. Otherwise, the advertiser with the maximum bid will win the ad impression.

6. AdX announces the result of auction to all DSPs, and send the final winner advertiser's ads to SSP.

7. SSP helps to display the ads to the available ad slots on the publisher's webpages in front of the users search results.

IV. KEY ISSUES IN BID OPTIMIZATION

Bidding optimization is one of the most crucial problems for maximizing the advertiser's profit in real time online advertising environment. The keyword basis bids are usually assumed to be static during the online auction process. However, the user queries sequence (containing impressions and auctions for online advertising) creates a complicated dynamic real time bidding environment where a real time bidding strategy and policies could significantly boost advertiser's profit. This became important on e-commerce websites auction platforms since impressions more readily convert into purchases. Here, analyzes the problem, Sponsored Search Real Time Bidding (SS-RTB), aims to generate proper bids on keyword basis at the impression level in the context of SS. Still, there is no publicly available solution to SS-RTB. Here, RTB problem studied in the context of display

advertising (DA). Nevertheless, SS-RTB is intrinsically different from traditional RTB. In DA, the impressions for bidding are concerned with ad position in publisher's web pages, while in SS the targets are ranking lists of dynamic keyword level user queries.

The main differences are:

(1) for a DA impression only the winning ad presented to the user otherwise not (i.e. a 0-1 problem), while in the Sponsored Search context multiple ads ranked high can be displayed to the query of the user;

(2) In Sponsored Search, adjust bid prices on multiple keywords for an ad to achieve optimal and efficient performance, while an ad in DA does not need such a keyword set. These differences develop popular methods for RTB in DA, such as predicting the winning market bidding price [15] or winning rate [16], which inapplicable in SS-RTB. Moreover, compared to ad placements in web pages, sequence of user query in SS are time variant and highly dynamic in nature. This creates for a complex model for SS-RTB.

Now, we will mathematically discuss the problem of bidding optimization in sponsored search Real time bidding auction platforms (SS-RTB).

In a simple and general scenario, an ad has a set of keyword tuples {$kwinf_1$, $kwinf_2$, . . ., $kwinf_n$}, where each tuple $kwinf_i$ can be defined as <belong_to, keyword, bid_price>. Typically, the bid_price here is predefined by the advertiser. The process of an auction *auct* could be explained as: every time a user(u) visits and types a query, the platform will retrieve a list of relevant matching keyword tuples, [$kwinf_1$, $kwinf_2$, . . . , $kwinf_r$][1], from the ad repository for auction. Each involved ad is then assigned a position ranking score according to its retrieved keyword tuple as bid_score $*$ bid_price. Here, bid_score is obtained from factors such as relevance, landing page experience, CTR and personalization, etc.

Finally, top ads will be presented to the user(u) and ranks accordingly. For SS-RTB, the main objective is to find optimal_bidprice rather than the bid_price for the keyword matched tuples, so as to maximize an ad's overall revenue. Here, concepts related to the e-commerce search scenario is used. Basically, this method is general and applicable to other search scenarios. Here, defining an ad's goal as maximizing the purchase amount PUR_AMT_d as income in a day d, while minimizing the cost $COST_d$ as expense in day d with a constraint that the PUR_AMT_d should not be lesser than the advertiser's expected value g.

$$\begin{aligned} \max \quad & PUR\_AMT_d / COST_d \\ s.t. \quad & PUR\_AMT_d >= g \end{aligned} \qquad (1)$$

Observing that PUR_AMT_d is highly similar with $COST_d$, we can change it into another equation:

$$\begin{aligned} max \quad & PUR\_AMT_d \\ s.t. \quad & COST_d = c \end{aligned} \qquad (2)$$

Eq. (2) is equivalent to Eq. (1) when $COST_d$ is highly related with PUR_AMT_d . The problem is to decide *opt_bidprice* in real-time scenario for an ad in terms of objective (2).

V. A SKETCH MODEL OF REINFORCEMENT LEARNING

Based on the problem we discussed in section 4, we now formulate it into a sketch model of RL:

State s: We used a general representation for states as $s = <b, t, auct>$, where *b* represents the budget left for the advertisement, *t* represents the step number of the decision sequence, and *auct* is the auction (impression) related feature vector that we can get from the advertising environment. It is worth to note that the b here is cost that the ad expects to expend in the left steps of auctions not the budget preset by the advertiser.

Action a: The decision of developing the real-time bidding price for each auction.

Reward *(s, a)*: The income under state s gained according to a specific action *a.*

Episode ep: one day is represent as an episode.

Finally, main goal is to find a policy P(s) which maps each states to an action *a*, to obtain the highest rewards: $\sum_{i=1}^{n} \gamma^{i-1} r(s_i, a_i)$. $\{\gamma^i\}$ is the set of discounted coefficients used in a standard Reinforcement Learning model [17].

Now, a series of regular periodic patterns observed and analyzed in real time data. By viewing the pattern, two different days share very similar dynamic patterns.

From Figure 2, which depicts the number of clicks through at different levels of aggregation (from seconds-level, minutes-level, hours-level) in January 28th, 2018 and January 29th, 2018 respectively. It can be analyzed that, the seconds-level curves do not display a similar pattern (Figure 2(a) and (b)), while from both minutes-level and hours-level we can analyzed a similar wave shape.

In addition, it also observed that the hours-level curves are more similar than minutes-level and seconds-level. We have similar analysis on other aggregated measures. The periodic patterns of the same measurements are very similar between 2 days at hour-level. By analysis of these regular patterns, we will take advantage of hours-level aggregated features rather than auction-level features to formulate the MDP model.Each day is taken as 24 steps of auction game, an episode of any day would always share the same general experiences. For example, it will meet an auction vary between 4:00 AM to 8:00 AM, while facing a heavy

competitive auction peak at around 9:00AM and a huge amount of impressions purchased at around 8:00 PM. Overriding the sketch Robust MDP model as follows.

State Transition: The auctions of an episode of one day will be grouped into m (m = 24) groups according to the timestamp of each hours basis. Each group contains many auctions in the corresponding period. A *state s* is re-defined as *<b, t, g>*, where *b* is the budget left, *t* is the specific time sequence, g denotes the feature vector containing aggregated statistical features of auctions in time sequence *t* , e.g. number-of-clicks, number of impression, cost, click-through-rate (CTR), pay per click (PPC), conversion rate (CVR) etc. In the following, it is observed that the state transition probabilities are consistent between two days.

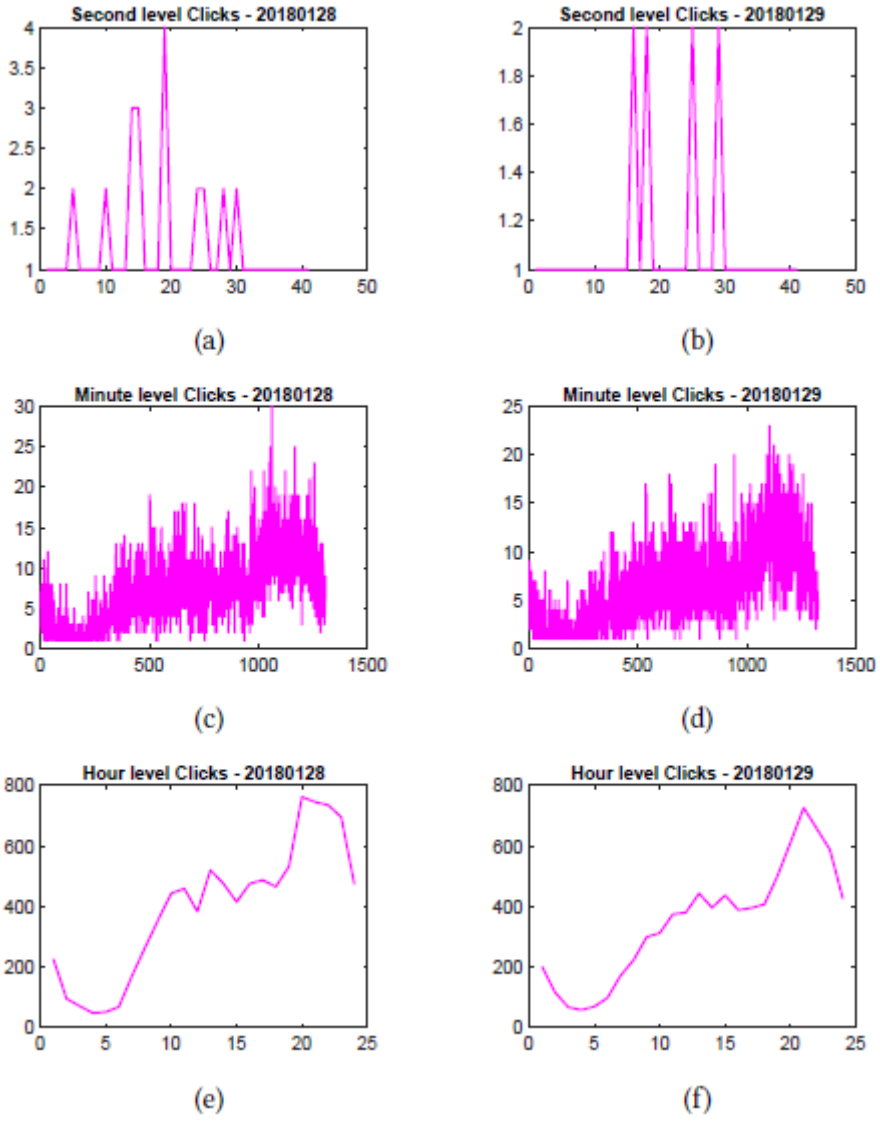

Figure 2: Patterns of Different Level Clicks

Action Space*:* Most Previous works used reinforcement learning (RL) to control the bid directly, so the main action is to set bid prices (costs). However, by applying the idea to our Robust MDP model would result in setting a group cost for all the auctions in the period or time sequence. It is very hard to derive impression level bid prices and this can't achieve real-time bidding. Instead of generating bid prices directly in dynamic environment, a control-by-model schemes used in which deployed a linear approximator function as the bidding model to fit the optimal bid_prices and utilize reinforcement learning (RL) to learn an optimal policy P(s) to control the real-time bidding model. Optimal bid price has a linear relationship with the CTR [19, 20]. Thus, an predicted conversion rate (PCVR) is defined as:

$$opt\_bidprice = f(PCVR) = \alpha \cdot PCVR \quad (3)$$

To sum up, the Robust MDP is modeled as follows:

State	$< b, t, g >$
Action	*set α*
Reward	*PUR_AMT gained in one step*
Episode	*a single day*

Table 1: Robust Model parameters

VI. EXPERIMENTAL ANALYSIS

Reinforcement Learning based MDP model is tested by online evaluation on a large e-commerce website of Alibaba Corporation with real advertisers and auctions in real time environment.
In this section, we discussed the dataset, compared methods and parameter setting used for evaluation.

6.1 Dataset

For online evaluation, randomly selected 1000 big ads in Alibaba's search real time auction platform, which nearly cover 100 million auctions per day. For online benchmark dataset is extracted from the search auction log for 2 days of the month of December, 2017. Each auction instance contains the bids, the clicks through, the auction ranking algorithms results and the corresponding predicted features i.e., PCVR. For evaluation, one day collection taken as the training data and the other day collection taken for testing over data. A/B test is performed over auction. Hence, observation of performance on the test collection. and the trained model is used to generate optimal real bidding decisions online.

6.2 Compared Methods and Evaluation Metric

The compared bidding methods in our experiments include:

- **Keyword-level bidding (KB):** It is based on keyword-level. It is an easy and efficient online approach adopted by search real time auction platforms. It is treated this as the fundamental baseline of the experiments.
- **RL with auction-level MDP (AMDP):** AMDP optimizes periodic bidding decisions by auction-level Deep Reinforcement Learning algorithm [18]. As in [18], this algo samples an auction in every 100 auctions interval as the next state.
- **RL with robust MDP (RMDP):** RMDP algorithms is used for bidding optimization which works well in dynamic environment. RMDP is based on single agent orientation without considering the competition between different ads.
- **Massive-agent RL with robust MDP (M-RMDP**): M-RMDP algorithm can be used for handling the massive-agent problem which is extension of RMDP.

For evaluation the performance of the algos, evaluation metric of *PUR_AMT /COST* under the same *COST* constraint is used. Here, used relative improved values with respect to KB. This doesn't affect performance comparison. Comparison is based for an official data exposure agreement currently.

6.3 Hyper-parameter Setting

Here, provide the settings of some key hyper-parameters. Same hyper-parameter setting and DNN network structure is used for all agents.

Target network update period	10000 steps
Memory size used	1 million
Learning rate	0.0001 with RMSProp method [26]
Batch size	300
Network structure	4-layer DNN
DNN layer size	[15, 300, 200, 100]

Table 2: Hyper-parameter Setting

VII. COMPARATIVE ANALYSIS

The results of online evaluation with a standard A/B testing configuration are observed. All the results are collected online and the key performance metric *PUR_AMT / COST* is used. Also showed results on metrics includes conversion rate (CVR), return-on-investment (ROI) and cost-per-click (PPC). These metrics are related to our observation.

7.1 Single-agent Analysis: The detailed performance of RMDP for each ad is listed in Table 3. It can be observed that the performance of RMDP out performs that of Keyword-level Bidding with an average of 35.04% improvement of *PUR_AMT /COST*. This suggests that RMDP model is indeed robust when deployed to real time auction platforms. Besides, we also considered performance metrics as a reference.
It is clearly observed that there is an average of 23.78 % improvement in conversion rate (CVR), an average of 21.38 % improvement in return-on-investment (ROI) and a slight average improvement (5.16%) in pay-per-click, PPC (i.e., the lower, the better). This means Robust MDP model indirectly improve other performance metrics that are generally considered in online advertising. The slight improvement in in pay-per-click (PPC) means that RMDP model helps advertisers save a little of his/her cost per click, although not prominent.

7.2 Multi-agent Analysis: A standard online A/B test is performed on the 1000 ads. The averaged improvement results of RMDP and M-RMDP are compared with Keyword-level Bidding are depicted in Table 4. It can be observed that M-RMDP outperforms the online Keyword-level Bidding algorithm and RMDP also in several aspects:
a) higher *PUR_AMT /COST* ratio, which is the main objective of our model

b) Higher ROI and CVR.

It is again observed that pay-per-click (PPC) is slightly improved, which means that this model can slightly help advertisers to save their cost per click (CPC). This relative performance improvement of RMDP is lower than that in the online single-agent case (see Table 3). The reason is that the competition between the ads affect its performance. In comparison, M-RMDP can well handle the multi-agent problem than RMDP.

Table 3: Performance improvement of RMDP compared to KB for single-agent case.

ad_id	PUR_AMT /COST	CVR	ROI	PPC
740053750	65.19%	60.78%	19.01%	-2.67%
749798178	4.15%	7.98%	0.63%	-11.66%
75694893	23.59%	8.83%	12.75%	-11.94%
781346990	41.6%	49.12%	43.86 %	5.33%
781625444	9.79%	30.95%	7.73%	14.28%
783136763	55.853%	27.03%	52.87%	-18.49%
750569395	2.854%	1.65%	19.60%	8.81%
787215770	21.52%	32.97%	46.94%	- 8.61%
805113454	57.08%	78.64%	68.73%	13.74%
802779226	31.44%	46.93%	19.97%	11.78%
Avg.	35.04%	23.11%	21.38%	-5.16%

Table 4: Performance improvement of RMDP compared to KB for multi-agent case.

Algorithm	PU R_AMT /COST	ROI	CVR	PPC
RMDP	6.29%	26.51%	3.12%	-3.36%
M-RMDP	13.01%	39.12%	12.62%	-0.74%

7.3 Convergence Analysis

Convergence analysis of the RMDP model is provided by two example ads in Figure 3. Figures 3(a) and 3(c) show the Loss curves with the number of learning grouped process. Figures 3(b) and 3(d) present the *PUR_AMT* (i.e. bid optimization objective in Eq. (2)) curves accordingly. We can observe that, in Figures 3(a) and 3(c) the Loss starts as or rapidly increases to a high value and then slowly converge to a smaller value, while in the same group range. *PUR_AMT* ratio improves persistently and becomes stable (Figures 3(b) and 3(d)). This provides us a good evidence that this algorithm has a solid capability to adjust bidding from a random policy to an optimal bidding policy solution. The curves in Figure 3 explained a good convergence performance of Reinforcement Learning (RL).

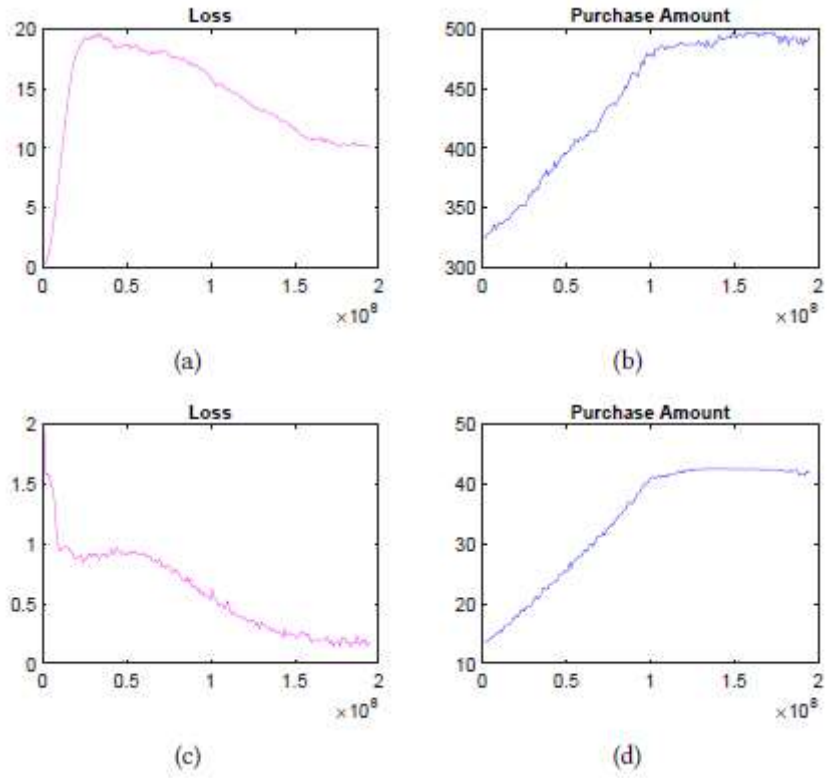

Figure 3: convergence Performance

VIII. CONCLUSION AND FUTURE SCOPE

Bidding Optimization is the critical problem in Online Advertising Environment in present time. Here, Reinforcement Learning is used in Robust Markov Decision Process (RMDP) model for optimization of bidding. It compared the data set for two days and training and testing is performed for evaluation of results. RMDP model is also compared with other bidding process based on the performance metric. It works well in performance improvement for single-agent case but for multi-agent case, its performance lacks than Massive-agent RL with robust MDP(M-RMDP). In comparison, M-RMDP can well handle the multi-agent problem than RMDP.

REFERENCES

[1] Yong Yuan, Feiyue Wang, Juan juan Li, Rui Qin. A Survey on Real Time Bidding Advertising. Published in IEEE transaction 978-1-4799-6058-3, 2014.
[2] Shuai Yuan, Jun Wang and Xiaoxue Zhao. Real-time Bidding for Online Advertising: Measurement and Analysis The Quarterly Journal of Economics, 108(4):941-964, 2013.
[3] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords," National Bureau of Economic Research, Tech. Rep., 2005.
[4] G. Aggarwal, A. Goel, and R. Motwani, "Truthful auctions for pricing search keywords," in EC. ACM, 2006.
[5] M. Cary, A. Das, B. Edelman, I. Giotis, K. Heimerl, A. R. Karlin, C. Mathieu, and M. Schwarz, in EC. ACM.
[6] N. Arnosti, M. Beck, and P. Milgrom, "Adverse selection and auction design for internet display advertising," Stanford University, Tech. Rep., 2014.
[7] Y. J. Hu, "Performance-based pricing models in online advertising," Social Science Research Network, 2004.
[8] Y. Moon and C. Kwon, "Online advertisement service pricing and an option contract," Electronic Commerce Research and Applications, vol. 10, no. 1, pp. 38–48, 2011.
[9] Jun Zhao, Guang Qiu, Ziyu Guan and Xiaofei He," Deep Reinforcement Learning for Sponsored Search Real-time Bidding" in ACM,2018.
[10] Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami,and Mohammad Mahdian. 2007. Dynamics of bid optimization in online advertisement auctions. In Proceedings of the 16th international conference on World Wide Web. ACM, 531–540.
[11] Kareem Amin, Michael Kearns, Peter Key, and Anton Schwaighofer. 2012. Budget optimization for sponsored search: Censored learning in MDPs. arXivpreprint arXiv:1210.4847 (2012).
[12] 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. ACM, 661–670.
[13] Jon Feldman, S Muthukrishnan, Martin Pal, and Cliff Stein. 2007. Budget optimization in search-based advertising auctions. In Proceedings of the 8th ACM conference on Electronic commerce. ACM, 40–49.

[14] Muthukrishnan,S .,A dX: a model for ad exchanges. ACM SIGecom Exchanges, 2009. 8(2): p. 9.
[15] Wush Chi-HsuanWu, Mi-Yen Yeh, and Ming-Syan Chen. 2015. Predicting winning price in real time bidding with censored data. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 1305–1314.
[16] Weinan Zhang, Shuai Yuan, and Jun Wang. 2014. Optimal real-time bidding for display advertising. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 1077–1086.
[17] Richard S Sutton and Andrew G Barto. 1998. Reinforcement learning: An introduction.
Vol. 1. MIT press Cambridge.
[18] Yu Wang, Jiayi Liu, Yuxiang Liu, Jun Hao, Yang He, Jinghe Hu, Weipeng Yan, and Mantian Li. 2017. LADDER: A Human-Level Bidding Agent for Large-Scale Real-Time Online Auctions. arXiv preprint arXiv:1708.05565 (2017).
[19] Kuang-chih Lee, Burkay Orten, Ali Dasdan, and Wentong Li. 2012. Estimating conversion rate in display advertising frompast erformance data. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 768–776.
[20] Claudia Perlich, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder, and Foster Provost. 2012. Bid optimizing and inventory scoring in targeted online advertising. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 804–812.