Optimizing the Segmentation Granularity for RTB Advertising Markets with a Two-stage Resale Model

Rui Qin, Yong Yuan*, Juanjuan Li
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Qingdao Academy of Intelligent Industries, Qingdao, China
Beijing Engineering Research Center of Intelligent Systems and Technology, Beijing, China
Email: rui.qin@ia.ac.cn, yong.yuan@ia.ac.cn (Corresponding author)
juanjuan.li@ia.ac.cn

Fei-Yue Wang, *Fellow*, *IEEE*
The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Qingdao Academy of Intelligent Industries, Qingdao, China
Research Center of Military Computational Experiments and Parallel System, National University of Defense Technology, Changsha, China
Email: feiyue.wang@ia.ac.cn

***Abstract*—Real Time Bidding (RTB) is an emerging business model and a popular research topic of online advertising markets. Using cookie-based big-data analysis, RTB advertising platforms have the ability to precisely identify the features and preferences of online users, segment them into various kinds of niche markets, and thus achieve the precision marketing via delivering advertisements to the best-matched users. The segmentation granularity used by such platforms, typically referred to as the Demand Side Platforms (DSPs), plays a central role in the effectiveness and efficiency of the RTB ecosystem. In practice, fine-grained user segmentations may lead to increased value-per-clicks and bid prices from advertisers, but at the same time reduced competition and possibly decreased bid prices in each niche market. This motivates our research on the optimal segmentation granularity to solve this dilemma faced by DSPs. Using a RTB market model with two-stage resales, we analyzed DSPs' segmentation strategies taking the revenues of both advertisers and DSPs into consideration. We also validated our proposed model and analysis using the computational experiment approach, and the experimental results indicate that with the increasing of segmentation granularity, the weighted sum of the DSP and advertisers' revenues tends to first rise and then decline in all weight-value cases, and the optimal granularity is greatly influenced by the value of weights. Our work highlights the need for DSPs of moderately using, instead of overusing, the online big data for maximized revenues.**

Keywords: real time bidding, demand side platforms, market segmentation, two-stage resale model, computational experiment

I. Introduction

With the rapid development and integration of Internet economy and big data analysis, Real Time Bidding (RTB) has emerged to be one of the most popular business model for online advertising and digital media [2, 4]. Using this RTB channel, advertisers can easily reach their target audiences via big-data-driven user profiling, and display their advertisements directly to the best-matched audiences with lowered costs. As such, RTB is widely recognized as an effective and efficient advertising format for online marketing.

The key advantage of RTB advertising is precision marketing, which is realized through the market segmentation strategies of Demand Side Platforms (DSPs). As a central part of the RTB ecosystem, DSPs have the potential of improving the match quality between advertisers and their target audiences through market segmentation, which can directly increase the effectiveness of RTB advertising. As such, market segmentation has been widely considered as the most important and challenging task for DSPs.

Market segmentation, which is also called user segmentation, has been widely studied in marketing research, and is regarded as the most effective way to improve the targeting accuracy for advertisers [15]. User segmentation aims to divide the users into multiple distinct groups via clustering methods, such as k-means and probabilistic-density-based mixture models [14]. In practice, a simple static feature such as age can be used to segment users into multiple groups with different age intervals [7], and such segmentation method is particularly effective for age-specific products and services [3, 6]. Generally, users' online behavior contains implicit signals about their interests and preferences [17], and thus many behavior-driven user segmentation approaches were proposed by analyzing the historical user activity [1]. As the most useful user activity, search behavior is regarded as an important indicator to reveal users' behavior pattern. Based on user queries, a topic-based user segmentation algorithm was proposed [12], which can divide the users with similar query terms into the same group. Moreover, a lot of hidden semantics maybe embedded in users' search behavior. As such, researchers have proposed some latent semantic user segmentation approaches based on latent Dirichlet allocation [5] or probabilistic latent semantic approach [16], so as to mine the hidden semantics and maximize the value of search behavior. Besides the search behavior, the evaluation behavior is another useful user activity, which reflects users' implicit preferences. By studying the semantic overlapping between the classes of items positively evaluated by users and the rest of classes, Saia et al. [11] proposed an interpretable and non-trivial user segmentation approach to uncover the implicit preferences, which can help advertisers find their desired target audiences.

As for RTB advertising, the research on market segmentation is still far from enough. In RTB practice, such distinguishable features as users' gender, age, interest and purchasing intention can be extracted as tags through cookie-based big-data analysis, and used for characterizing online users. Each tag can divide the users into multiple audience groups. Typically, different choices and combinations of tags may result in a different market segmentation granularity, which is a key parameter to determine both the competition degree among advertisers and their valuation of ad impressions. On one hand, fine-grained market segmentation typically leads to increased advertiser-audience match quality, and in turn increased value-per-clicks and bid prices from advertisers. On the other hand, however, fine-grained market segmentation will also decrease the number of advertisers in each niche market, resulting in reduced competition and possibly lowered bid prices in each niche market [8]. As such, there is a critical need for DSPs to determine the optimal granularity of market segmentation so as to maximize their revenues and improve the effectiveness of the RTB ecosystem.

In this paper, we strive to optimize the granularity of market segmentation using a two-stage resale market model that is commonly used in RTB practice. We take the revenues of both DSPs and advertisers into consideration, and propose a mathematical programming model for optimizing the market segmentation granularity to maximize the weighted sum of their revenues. Considering the theoretical intractability of the proposed model, we utilize the computational experiment approach to validate our model. The experimental results show that the market segmentation granularity has great influence on the weighted-sum revenues of both advertisers and DSPs. With the increasing refinement of market segmentation granularity, the total revenue has a tendency of a rise first and followed by a decline, in all weight-value cases. Furthermore, the optimal strategies for market segmentation differs in terms of the values of weights, which represent the allocation of maximized revenues between the DSP and advertisers.

The remainder of this paper is organized as follows. In Section II, we introduce the two-stage resale model of RTB advertising, briefly state our research problem, and propose our research model on market segmentation and its solution procedure. In Section III, we design numerical experiments using the computational experiment approach to validate our model and analysis. Section IV discusses the managerial insights of our research findings, and Section V concludes our research efforts.

II. Model of Market Segmentation

A. Two-stage Resale Model

In RTB advertising, each ad request will trigger an auction session with a two-stage bidding process, in which DSPs serve as resellers in pursuit of intermediate fees in each auction.

Figure 1 presents the two-stage resale process of an ad request in a RTB system with n DSPs. Once an ad request arrives, each DSP will identify the interests and characteristics of the user behind the ad request, and start the first-stage auction asking for bids from all eligible matched advertisers

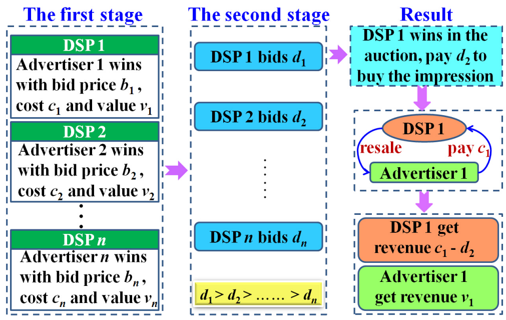

Fig. 1. The two-stage resale process in RTB advertising

registered on it. The winning advertiser on each DSP i, together with his/her bid b_i, cost c_i and value-per-click v_i, will enter the second-stage auction, where each DSP i will submit a bid d_i according to its winning bid b_i. For instance, if $d_1 > d_2 > \cdots > d_n$, then DSP 1 wins the second-stage auction and the ad impression with the cost d_2. When reselling this ad impression back to its winning advertiser, the advertiser needs to pay c_1 to DSP 1 and get revenue v_1, while DSP 1 needs to pay d_2 and get revenue $c_1 - d_2$ from the ad impression.

B. Problem Statement

In RTB markets, various kinds of tags can be extracted from cookie data to characterize the target audiences (or users), and each tag can divide the audiences into multiple groups. Thus, a hierarchical structure of the users can be constructed by increasing and refining tags. Figure 2 presents a $(M+1)$-layer structure of the users including M tags, where the ith tag can divide the users into t_i groups. Obviously, the lower layers contains more tags, resulting in better matching quality and less users in each niche market.

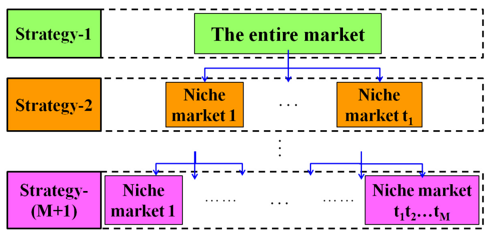

Fig. 2. The user structure

For DSPs, each layer corresponds to one of their segmentation strategies, which lead to different marketing effects for advertisers and different revenues for DSPs. Typically, the optimal strategy that maximizes the DSP's revenue differs from that maximizes advertisers' revenues. This makes our strategy optimization fall into the classical principle-agent problem in information economics. Without loss of generality, in our research, we will investigate the optimal segmentation strategy that maximizes the weighted sum of revenue of both the DSP and its advertisers.

C. Notations

Suppose there are M alternative tags for characterizing the target audiences, denoted by $T = \{T_1, T_2, \cdots, T_M\}$, and the audiences can be segmented to t_i groups with tag T_i. Using these tags, DSP can segment the RTB market with $M+1$ diverse granularities, represented by $L = \{L_1, L_2, \cdots, L_{M+1}\}$, and L_i is generated with $i-1$ tags. There are $\Pi_{j\leq i-1}t_j$ niche markets for L_i, where $t_0=1$. Generally, each granularity corresponds to one of the DSP's segmentation strategy. As such, the DSP has totally $M+1$ market segmentation strategies.

Suppose there are S ad requests in a given time period, denoted as $Q=\{q_1,q_2,\cdots,q_S\}$. The reserve price of each ad request $q_j\in Q$ is denoted as ρ_j . Denote the advertiser set on the DSP as $U=\{u_1,u_2,\cdots,u_N\}$, and the corresponding set of advertisers' budgets as $B=\{B_1,B_2,\cdots,B_N\}$.

For simplicity, we can reasonably assume that there is only one niche market matched with each advertiser, and thus each ad request falls into only one niche market under a specific granularity. If we represent the matching probability between the advertiser $u_k\in U$ and the ad request $q_j\in Q$ under granularity L_i as $\sigma_i(k,j)\in\{0,1\}$, then $\sigma_i(k,j)=1$ if and only if ad impression q_j falls into the niche market matched with advertiser u_k under granularity L_i. For convenience, we can also characterize the targeting of the advertisers with the M tags. In this case, if each tag in $\{T_1,T_2,\cdots,T_{i-1}\}$ for advertiser u_k and ad request q_j takes the same value under granularity L_i, then $\sigma_i(k,j)=1$. Generally, the matching probability is used to measure the matching degree of the ad impression with the advertiser, and only if $\sigma_i(k,j)=1$, the advertiser u_k will participate in the auction and bid for q_j.

Let $v_i(k,j)$ be the value function of the advertiser $u_k\in U$ for ad impression q_j under granularity L_i. Then we can assume that the bid of advertiser u_k for ad impression q_j is also $v_i(k,j)$ according to the equilibrium outcome of Vickrey auction mechanism [8].

D. Market Segmentation Model

Suppose there are K competing DSPs in the RTB market, and the winning advertiser in one DSP can obtain the ad impression only if he/she beats all the winning advertisers on other DSPs.

Under granularity L_i, the advertisers on the DSP bidding for q_j can be given as follows

$$U_i(j)=\{u_k\in U|\sigma_i(k,j)=1,b_i(k,j-1)\geq v_i(k,j)\}, \quad (1)$$

in which the advertisers who bid the highest and the second highest can be found by

$$\begin{aligned}u_{k^*(i,j)}&=\text{argmax}_{u_k\in U_i(j)}v_i(k,j),\\ u_{k'(i,j)}&=\text{argmax}_{u_k\in U_i(j)/u_{k^*(i,j)}}v_i(k,j).\end{aligned} \quad (2)$$

Obviously, $u_{k^*(i,j)}$ wins in the DSP.

Suppose the bid of the DSP for ad request q_j is $d_1(i,j)$, and the highest bid of all the other DSPs is $d_2(i,j)$. Then we can use the following indicator function to define whether or not the DSP wins ad request q_j under granularity L_i:

$$I(i,j)=\begin{cases}1, & \text{if } d_1(i,j)>d_2(i,j)\\ 0, & \text{otherwise.}\end{cases} \quad (3)$$

Thus, if $I(i,j)=1$, the DSP wins the second-stage auction, and the winning advertiser on it obtains the ad impression. According to the RTB auction mechanism, the advertiser with the highest bid wins the auction and pays the second highest bid [9]. Thus, advertiser $u_{k^*(i,j)}$ needs to pay

$$c_i(k^*(i,j),j)=\max\{v_i(k^{'}(i,j),j),\rho_j\}I(i,j) \quad (4)$$

to the DSP, and the DSP needs to pay

$$c_1(i,j)=\max\{d_2(i,j),\rho_j\}I(i,j) \quad (5)$$

for the ad impression.

Assume the revenue of the advertiser from an ad impression is equal to his/her value for the impression, then the winning advertiser $u_{k^*(i,j)}$ obtains revenue $v_i(k^*(i,j),j)I(i,j)$ from q_j, and the remaining budget for advertiser u_k after q_j is

$$b_i(k,j)=\begin{cases}b_i(k,j-1)-c_i(k^*(i,j),j), & \text{if } k=k^*(i,j)\\ b_i(k,j-1), & \text{otherwise,}\end{cases} \quad (6)$$

where $b_i(k,0)=B_k$. Besides, the revenue of the DSP from ad impression q_j is $c_i(k^*(i,j),j)-c_1(i,j)$.

Denote the revenue of the DSP and all the advertisers from ad impression q_j under the granularity L_i as $r_1(i,j)$ and $r_2(i,j)$, respectively. Then we have

$$r_1(i,j)=c_i(k^*(i,j),j)-c_1(i,j), \quad (7)$$

$$r_2(i,j)=v_i(k^*(i,j),j)I(i,j). \quad (8)$$

We assume that DSPs are rational and accept only nonnegative revenues from auctions. It is worth noting that in many cases, such as $d_1(i,j)>d_2(i,j)>v_i(k^{'}(i,j),j)>\rho_j$, the DSP wins with negative revenue $v_i(k^{'}(i,j),j)-d_2(i,j)<0$. In order to eliminate such cases, the DSP's bid $d_1(i,j)$ should be no larger than $v_i(k^{'}(i,j),j)$. In our model, we assume $d_1(i,j)=v_i(k^{'}(i,j),j)$ for simplicity.

As such, under granularity L_i, the DSP's revenue and advertisers' total revenue generated from all ad requests $q_j\in Q$ can be obtained by

$$g_1(i)=\textstyle\sum_{j\in Q}r_1(i,j), g_2(i)=\sum_{j\in Q}r_2(i,j). \quad (9)$$

Typically, $g_1(i)$ and $g_2(i)$ will not reach their maximums under the same granularity L_i. Thus, the DSP has to take both its own revenue and the advertisers' total revenue into consideration through the following formula

$$g(i)=f(g_1(i),g_2(i))=wg_1(i)+(1-w)g_2(i), \quad (10)$$

where $w\in[0,1]$ is the weight of the DSP's revenue in the objective function.

With the above analysis, we can formulate our market segmentation model as follows:

$$\begin{cases}\max\limits_{L_i\in L} g(i)=wg_1(i)+(1-w)g_2(i)\\ \text{subject to:}\\ U_i(j)=\{u_k|u_k\in U,\sigma_i(k,j)=1,b_i(k,j)\geq v_i(k,j)\}\\ u_{k^*(i,j)}=\text{argmax}_{u_k\in U_i(j)}v_i(k,j)\\ u_{k_i^{'}(j)}=\text{argmax}_{u_k\in U_i(j)/u_{k^*(i,j)}}v_i(l,j)\\ I(i,j)=\begin{cases}1, & \text{if } v_i(k^{'}(i,j),j)>d_2(i,j)\\ 0, & \text{otherwise}\end{cases}\\ c_i(k^*(i,j),j)=\max\{v_i(k_i^{'}(j),j),v^{''}(j),\rho_j\}I(i,j)\\ c_1(i,j)=\max\{d_2(i,j),\rho_j\}I(i,j)\\ r_1(i,j)=v_i(k^*(i,j),j)I(i,j)\\ r_2(i,j)=c_i(k^*(i,j),j)-c_1(i,j)\\ g_1(i)=\sum_{j\in Q}r_1(i,j)\\ g_2(i)=\sum_{j\in Q}r_2(i,j)\\ b_i(k,j)=\begin{cases}b_i(k,j-1)-c_i(k^*(i,j),j), & \text{if } k=k^*(i,j)\\ b_i(k,j-1), & \text{otherwise}\end{cases}\\ b_i(k,0)=B_k\\ \text{for } i=1,2,\cdots,M+1, j=1,2,\cdots,S.\end{cases} \quad (11)$$

For any given w, we can determine the optimal granularity L_{i^*} and the corresponding optimal revenue $g(i^*)$ through solving this model (11). Besides, the DSP's revenue $g_1(i^*)$ and the advertisers' revenue $g_2(i^*)$ can also be determined, respectively.

E. The Solution

As can be seen in our model (11), there are lots of complex iterative processes for each granularity and each ad request. Moreover, the iterative processes are highly dependent between each ad request and its previous one. As such, it is rather difficult to derive its accurate numerical solution.

Our solution can be depicted as follows. In order to identify the optimal granularity, we first compute the weighted sum of the DSP and advertisers' revenues from each ad request under each granularity, and then the weighted total revenue under each granularity by summing up the revenues from all the ad requests. Finally, by comparing the weighted total revenues under all possible granularities, we can identify the optimal revenue and the corresponding granularity. The detailed process of our solution to model (11) can be described as follows:

Step 1 : Find the set of advertisers $U_i(j)$ on the DSP bidding for each ad request q_j under each granularity L_i using formula (1).

Step 2 : Find the advertisers $u_{k^*(i,j)}$ and $u_{k'(i,j)}$ with the highest and the second highest bids on the DSP for ad request q_j from the advertiser set $U_i(j)$, according to formula (2).

Step 3 : Check if the DSP wins ad request q_j by comparing its bid $d_1(i,j)$ with the highest bid $d_2(i,j)$ of all the other DSPs, according to formula (3).

Step 4 : Compute the cost $c_i(k^*(i,j),j)$ of advertiser $u_{k^*(i,j)}$ and the cost $c_1(i,j)$ of the DSP for ad request q_j according to formula (4) and (5), respectively, and update the remaining budget $b_i(k,j)$ of each advertiser u_k according to formula (6).

Step 5 : Compute the revenues of the DSP and the advertisers from ad request q_j under granularity L_i according to formula (7) and (8), respectively.

Step 6 : Determine the total revenues $g_1(i)$ and $g_2(i)$ of the DSP and the advertisers by summing up the revenues from all ad requests $q_j \in Q$ according to formula (9).

Step 7 : Compute the weighted total revenue $g(i)$ for any given weight $w \in [0,1]$, according to formula (10).

Step 8 : Identify the maximized weighted total revenue $g(i^*)$ and the corresponding optimal granularity L_{i^*} according to formula (11).

III. Computational Experiments

Due to the essential model intractability and the lack of high-quality data in online RTB markets, it is quite difficult to validate our proposed model with online field experiments. Fortunately, with the help of computational experiments approach [13], we can design experiment scenarios to evaluate our model and analysis. In this section, we will utilize the computational experiment approach to validate the effect of market segmentation on the total revenue, and identify the optimal segmentation granularity for maximizing the advertisers' and the DSP's revenues.

A. Computational Experiment Scenario

We consider a randomly generated experiment scenario that there are 2 DSPs in the market, and they have the same winning probability in each auction. By analyzing the big data of target audiences, 9 tags are extracted to characterize the ad requests, and each tag can divide these ad requests into 2 groups. Utilizing these tags, 10 strategies can be used for the DSP to segment the market, with the number of tags as 0 to 9, respectively, as shown in Figure 3. For these strategies, the numbers of niche markets are $2^0 = 1, 2^1 = 2, 2^2 = 4, \cdots, 2^9 = 512$, respectively.

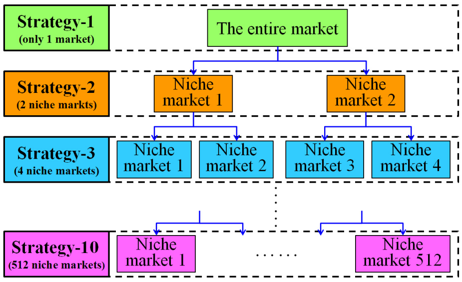

Fig. 3. The 10 strategies for segmenting the market and the corresponding niche markets

To evaluate these 10 strategies, we construct a computational experiment with randomly generated parameters of 2 DSPs, 1,000,000 ad requests and 100 advertisers. The ad requests are randomly distributed in these niche markets, and the total budgets of the advertisers are uniformly distributed in $[200, 1500]$. Figure 4 generates the above data in our experiment.

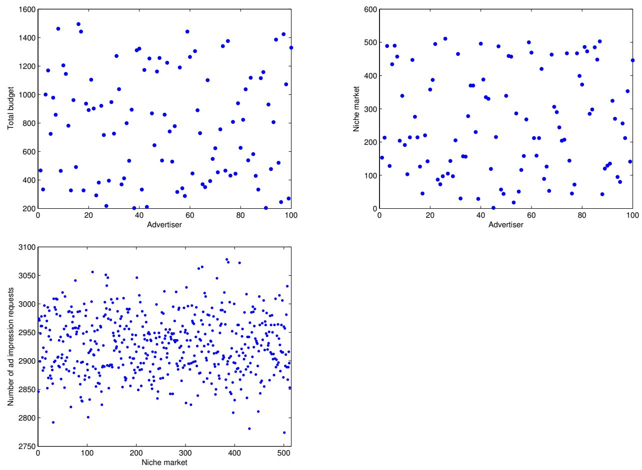

Fig. 4. The total budget of each advertiser, the target niche market of each advertiser and the total number of ad requests in each niche market under Strategy-10

Due to the positive correlations between the advertisers' values of ad impressions and the accuracy of matching [10], it is reasonable to assume that the CPMs (Cost Per Mille) of the advertisers increase with the granularity. Suppose the CPMs of the advertisers under Strategy-1 is uniformly distributed in $[2.00, 5.00]$, the lower bound and the upper bound will increase 1.00 and 2.00 each time, respectively, with the increasing of the granularity. Moreover, the advertiser with a higher CPM under one strategy is assumed to have a higher CPM under other strategies.

B. Experimental Results

In this section, we conduct 1000 independent computational experiments, aiming to obtain general conclusions for the optimal strategy under different weights ($w = 0, 0.1, 0.2, \cdots, 1$).

We first study the cases of $w = 0$ and $w = 1$. In both the two cases, the objective function can be simplified into optimizing the total revenue of the DSP (i.e., $w = 1$) or the advertisers (i.e., $w = 0$). The experimental results are provided in Figure 5, from which we can obtain the following results:

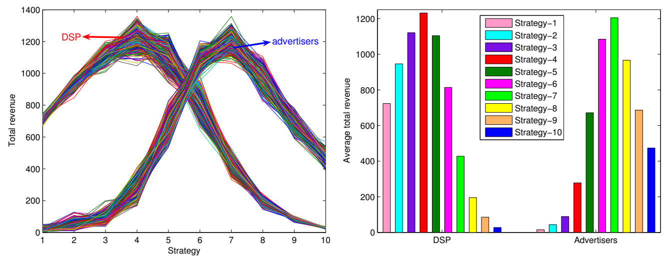

Fig. 5. Comparisons of the total revenues and the average total revenues for the DSP and the advertisers in the 1000 experiments

(1) In all the 1000 experiments, the total revenues of both the DSP and the advertisers have a tendency of a rise first and followed by a decline, with the increasing of granularity. The maximum occurs at Strategy-4 952 times out of 1000 times (95.2%) for the DSP, and Strategy-7 988 times out of 1000 times (98.8%) for the advertisers.

(2) The average total revenues of both the DSP and the advertisers also have a tendency of a rise first and followed by a decline, with the increasing of granularity. The maximum occurs at Strategy-4 for the DSP, and Strategy-7 for the advertisers.

Moreover, we study the effects of different weights ($w \in \{0, 0.1, 0.2, \cdots, 1\}$) on the weighted total revenues for these strategies, and the variations of the weighted total revenues under different strategies and different weights are given in Figure 6. Furthermore, by fixing one parameter, we can verify the effects of different strategies and w on the revenues, and the results are shown in Figure 7–Figure 8, respectively. The optimal strategies and the corresponding revenues of the DSP and the advertisers under different w are given in Table I. From these results, we can obtain the following conclusions:

(1) For all possible values of w, the weighted total revenues have a tendency of a rise first and followed by a decline, with the increasing of granularity.

(2) The optimal granularity may differ with the changes of the w values. Specifically, the optima occurs at Strategy-4 for $w = 0.8, 0.9, 1$, Strategy-5 for $w = 0.6, 0.7$, Strategy-6 for $w = 0.3, 0.4, 0.5$, and Strategy-7 for $w = 0, 0.1, 0.2$.

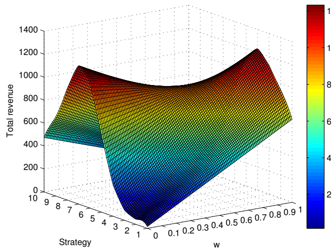

Fig. 6. Variations of the weighted total revenues under different strategies and w

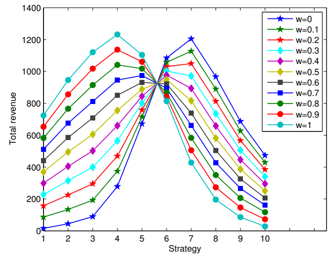

Fig. 7. The effect of segmentation strategy on the weighted total revenues

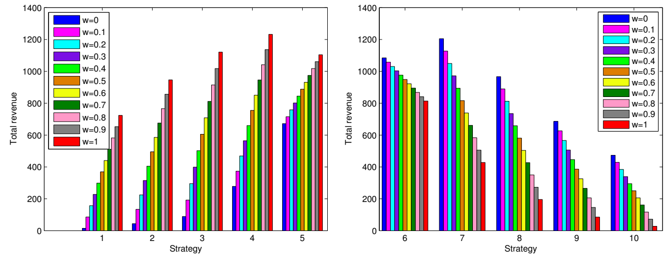

Fig. 8. Comparisons of the effect of w on the weighted total revenues under different segmentation strategies

TABLE I
THE OPTIMAL STRATEGY AND THE CORRESPONDING REVENUES OF THE DSP AND THE ADVERTISERS UNDER DIFFERENT w

w	Optimal strategy	The corresponding revenue	
		DSP	Advertisers
0, 0.1 0.2	Strategy-7	428.17281	1204.848
0.3,0.4,0.5	Strategy-6	813.72377	1084.670
0.6,0.7	Strategy-5	1104.20787	671.925
0.8,0.9,1.0	Strategy-4	1231.93101	278.428

(3) For the DSP, the optimal strategy is to choose the value of w in $\{0.8, 0.9, 1.0\}$ in our experiment and set the granularity as in Strategy-4; for the advertisers, the optimal strategy is to choose the value of w in $\{0, 0.1, 0.2\}$ and set the granularity as in Strategy-7.

(4) For different strategies, the effect of w on the weighted total revenues may be entirely different. Specially, for Strategy-1 – Strategy-5, the weighted total revenues are monotone increasing with w, while monotone decreasing for Strategy-6 – Strategy-10.

C. Result Analysis

From the experimental results in the above section, we can obtain the following research findings:

(1) The weighted total revenue will not always increase with the increasing refinement of granularity. There exists a threshold, before which the revenues can be improved by increasingly fine-grained segmentation while decreasing sharply in case that the granularity exceeds the threshold. Furthermore, this threshold will be greatly influenced by the weight values (w), and different choices of the weights will lead to different thresholds.

(2) Larger w indicates that the DSP pays more attention to improving its own revenue than the advertisers' revenues. Thus, the deduced optimal granularity is more favorable to the DSP instead of the advertisers, and vice versa.

IV. Managerial Insights

Our research findings can offer useful managerial insights for DSPs' decision making in RTB advertising markets. On one hand, big-data-driven precision marketing is widely considered as one key advantage of RTB advertising. Intuitively, one might expect that big data will lead to better revenues for DSPs. However, our study offers a counterintuitive conclusion that, although online big data has the potential of improving the RTB advertising effectiveness via increased advertiser-audience match quality, it does not always result in better revenues. Instead, we prove that there exists an optimal granularity for RTB market segmentation, beyond which the revenues will decrease due to reduced competition among advertisers in RTB auctions. As such, our work can be considered as a counter-example to the "bigger data is better" idea in RTB advertising, and also highlights the need for DSPs of moderately using the online big data for maximized revenues with optimal "resolutions" in user profiling.

On the other hand, our study offers an actionable solution framework for DSPs' market segmentation, especially when faced with the principal-agent decision-making scenario. In RTB practice, most DSPs aim at maximizing their own revenues without consideration of advertisers' revenues. Our work indicates that the weighted sum of the DSP and advertisers' revenues can serve as a good optimization objective to derive an effective solution that benefits both the DSPs and advertisers. By dynamically adjust the weighted allocation parameter w, DSPs can easily derive the optimal granularity that corresponds to various combinations of the total revenues.

V. Conclusions and Future Work

Market segmentation is an important task for DSPs, and plays a critical role in maintaining the effectiveness and efficiency of RTB advertising. In this paper, we established a RTB advertising model with two-stage resales, and optimized the market segmentation granularity so as to maximize the weighted sum of the DSP and advertisers' revenues. The computational experiment approach is used to evaluate our model and solution. In our future work, we are planning to extend this paper from the following aspects: (a) Studying the principal-agent games played by advertisers and the DSPs, and analyze their bidding behavior and the resulting Nash equilibrium continuum; (b) Comparing the market segmentation strategies under all exising profit models adopted by DSPs, and exploring new profit models for DSPs.

Acknowledgment

This work is partially supported by NSFC (71472174, 71232006, 61533019, 61233001, 71402178) and the Early Career Development Award of SKLMCCS (Y6S9011F4E).

References

[1] J. Bian, A. Dong, X. He, S. Reddy, and Y. Chang, "User action interpretation for online content optimization", IEEE Transactions on Knowledge and Data Engineering, 25(9): 2161–2174, 2013.

[2] R. Cavallo, P. McAfee, and S. Vassilvitskii, "Display advertising auctions with arbitrage", ACM Transactions on Economics and Computation, 3(3): 15, 2015.

[3] S. Faggian, and L. Grosset, "Optimal advertising strategies with age-structured goodwill", Mathematical Methods of Operations Research, 78(2): 259–284, 2013.

[4] J. Feldman, V.S. Mirrokni, S. Muthukrishnan, and et al., "Auctions with intermediaries", Proceedings of the 11th ACM Conference on Electronic Commerce, 2010, pp. 23–32.

[5] X. Gong, X. Guo, R. Zhang, X. He, and A. Zhou, "Search behavior based latent semantic user segmentation for advertising targeting", Proceedings of the 13th IEEE International Conference on Data Mining, 2013, pp. 211–220.

[6] L. Grosset, and B. Viscolani, "Advertising for the introduction of an age-sensitive product", Optimal Control Applications and Methods, 26(3): 157–167, 2005.

[7] J. Mazanee, "Market Segmentation", in Encyclopedia Tourism, London: Routledge, 2000.

[8] R. Myerson, "Optimal auction design", Mathematics of Operations Research, 6(1): 58–73, 1981.

[9] S. Muthukrishnan, "Ad exchanges: Research issues", Internet and Network Economics, 1–12, 2009.

[10] M. Mobius, H. Nazerzadeh, G. Lewis, and et al., "Buy-it-now or take-a-chance: A new pricing mechanism for online advertising", Society for Economic Dynamics, paper 443, 2012.

[11] R. Saia, L. Boratto, and S. Carta, "A latent semantic pattern recognition strategy for an untrivial targeted advertising", Proceedings of the 4th International Congress on Big Data, 2015, pp.491–498.

[12] S. Tu, and C. Lu, "Topic-based user segmentation for online advertising with latent dirichlet allocation", Proceedings of the 6th International Conference on Advanced Data Mining and Applications, 2010, pp. 259–269.

[13] F.Y. Wang, "Artificial societies, computational experiments, and parallel systems: A discussion on computational theory of complex social-economic systems", Complex Systems and Complexity Science, 1(4): 25–35, 2004.

[14] M. Wedel, and W.A. Kamakura, "Market Segmentation: Conceptual and Methodological Foundations", Springer Science & Business Media, 2012.

[15] A. Weinstein, "Handbook of Market Segmentation: Strategic Targeting for Business and Technology Firms", Routledge, 2013.

[16] X. Wu, J. Yan, N. Liu, S. Yan, Y. Chen, and Z. Chen, "Probabilistic latent semantic user segmentation for behavioral targeted advertising", Proceedings of the 3rd International Workshop on Data Mining and Audience Intelligence for Advertising, 2009, pp. 10–17.

[17] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, "How much can behavioral targeting help online advertising?", Proceedings of the 18th International Conference on World Wide Web, 2009, pp. 261–270.