Exploring the optimal granularity for market segmentation in RTB advertising via computational experiment approach

Rui Qin a,b,d, Yong Yuan a,b,⇑, Fei-Yue Wang a,b,c

a The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
b Institute of Parallel Economics, Qingdao Academy of Intelligent Industries, Qingdao, China
c Research Center of Military Computational Experiments and Parallel System, National University of Defense Technology, Changsha, China
d Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China

A R T I C L E   I N F O

Article history:
Received 26 January 2017
Received in revised form 28 May 2017
Accepted 2 July 2017
Available online 5 July 2017

Keywords:
Real Time Bidding
Demand Side Platforms
Market segmentation
Computational experiment

A B S T R A C T

Real Time Bidding (RTB) is a novel business model of online computational advertising, developing rapidly with the integration of Internet economy and big data analysis. It evolves the business logic of online ad-delivery from buying “ad-impressions” in websites or ad slots to directly buying the best-matched “target audiences”, and thus can help advertisers achieve the precision marketing. As a critical part of RTB advertising markets, Demand Side Platforms (DSPs) play a central role in matching advertisers with their target audiences via cookie-based data analysis and market segmentation, and their segmentation strategies (especially the choice of granularity) have key influences in improving the effectiveness and efficiency of RTB advertising markets. Based on a mathematical programming approach, this paper studied DSPs’ strategies for market segmentation, and established a selection model of the granularity for segmenting RTB advertising markets. With the computational experiment approach, we designed three experimental scenarios to validate our proposed model, and the experimental results show that: 1) market segmentation has the potential of improving the total revenue of all the advertisers; 2) with the increasing refinement of the market segmentation granularity, the total revenue has a tendency of a rise first and followed by a decline; 3) the optimal granularity of market segmentation will be significantly influenced by the number of advertisers on the DSP, but less influenced by the number of ad requests. Our findings show the crucial role of market segmentation on the RTB advertising effect, and indicate that the DSPs should adjust their market segmentation strategies according to their total number of advertisers. Our findings also highlight the importance of advertisers as well as the characteristics of the target audiences to DSPs’ market segmentation decisions.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

With the rapid development and integration of Internet economy and big data analysis, Real Time Bidding (RTB) has been widely-recognized as the main-stream business model of online advertising markets (Cavalló et al., 2014; Feldman et al., 2010), and an inevitable evolving trend of the sales model for most online digital media in the near future. RTB can help advertisers reach their target audiences via big-data-driven real-time matching and auction-based dynamic pricing, and thus can display the advertisements to the right audiences at the right time with lowered costs. As such, RTB has the potential of improving the market efficiency as well as advertisers’ revenue.

As a central part in the RTB ecosystems, Demand Side Platforms (DSPs) serve as agencies making decisions on behalf of their advertisers (i.e., the demand side). Typically, the key decisions for DSPs include designing efficient bidding algorithms, and formulating effective strategies for market segmentation. In literature, designing bidding algorithms has attracted intensive interests from researchers. For instance, Ghosh et al. (2009) proposed an offline bidding algorithm for DSPs that is suitable for both the full information and the partially observable information settings. Other offline algorithms have been proved to be able to optimize advertisers’ bids based on the historical average winning bid prices (Zhang and Zhang, 2014), the predicted winning rates and prices (Li and Guan, 2014), as well as the click-through-rates (Zhang et al., 2014). Chen et al. (2011) designed an online bidding algo-
rithm that can support fine-grained impression valuation and adjust value-based bids dynamically.

As for DSPs’ market segmentation strategy, however, related research is still far from enough. Via analyzing cookie-based online big data, such characteristics as gender, age, preference and purchasing intention can be elicited to profile the target audiences. Thus, each target audience will be “labeled” with lots of tags indicating his/her potential interest or preference to specific goods, services, and even advertisements. For example, the combination of tags “gender=’male’ & age=’20–25’ & interest=’sports’” is most likely to be a potential target audience of the advertiser owning a sports shop. With the help of these tags, DSPs can in theory infinitely precisely segment advertisers’ target audiences into large numbers of niche markets, resulting in better matching between the advertisers and their target audiences. In RTB practice, the granularity used by DSPs to segment their target audiences is proved to have great effect on the advertising effect of RTB advertising. For example, a sportswear firm can get a higher conversion rate when displaying their advertisements to the sport fans than the general audiences of the website (Mobius et al., 2012).

Actually, each tag can be used to divide the audiences into multiple groups, and the set of tags designed by a DSP determines its granularity for market segmentation. Generally speaking, the market segmentation granularity is a key parameter that can partly determine both the competition degree among advertisers and their valuations of ad impressions. On one hand, a fine-grained market segmentation can improve the precision of impression valuation and also advertisers’ value-per-ad-impression in each niche market. This helps increase the average price of RTB ads. On the other hand, however, with the increased segmentation granularity, the number of advertisers in each niche market, and also the resulting competition among them, will be reduced. This will decrease the RTB ad prices according to the classical auction theory (Myerson, 1981). For example, an empirical study shows that after Microsoft’s ADECN platform segments its market into large numbers of niches with only 2–3 advertisers, the advertising performance is witnessed to drop significantly due to the reduced competition (Bergemann and Bonatti, 2011; Levin and Milgrom, 2010). Obviously, there exists a dilemma for DSPs in choosing the granularity of market segmentation in pursuit of better ad prices and revenue. Therefore, it is of great theoretical and practical importance to study the market segmentation problems, and provide a feasible market segmentation strategy for DSPs, so as to maximize the marketing effect for advertisers in RTB markets.

Several challenges can be expected in studying DSPs’ segmentation strategy. First, due to the large number of heterogeneous advertisers served by a DSP, it is typically difficult to measure the advertising effect and find the optimal segmentation strategy for all advertisers. Second, there are large numbers of alternative tags to characterize the target audiences, and evaluating the effect of these tags is a critical but challenging issue. Third, it is naturally difficult to conduct repeated online experiments to validate the effectiveness of the research model in the real-time, dynamic and complex RTB markets.

In this paper, we aim to study the market segmentation problem for DSPs. We define this issue as an optimization problem of seeking for the optimal market segmentation granularity, and establish a mathematical programming model to solve this problem. Considering the market complexity, we utilize a computational experiment approach, and design experiments to validate the effectiveness of our proposed model. The experimental results show that the market segmentation granularity has great influence to the total revenue of all the advertisers. With the increasing refinement of the market segmentation granularity, the total revenue has a tendency of a rise first and followed by a decline. Furthermore, the optimal market segmentation granularity can be influenced by the number of advertisers on the demand side, but less influenced by the number of ad requests on the supply side.

It is worth noting that, one might expect that there are possibly two ways in studying the market segmentation issue in RTB markets. The first one is to regard the market segmentation granularity as an endogenous factor, and focus on its intrinsic influence on the advertising effect and revenue from a single-objective optimization viewpoint. The other one is to study the interaction and its equilibrium outcome between one DSP’s market segmentation decisions and other exogenous factors (e.g., other DSPs’ market segmentation strategies), typically from a game-theoretic viewpoint. In RTB practice, however, market segmentation strategies and the related tag taxonomy are always kept as DSPs’ top business secret and cannot be observed in markets, so that DSPs can hardly interact game-theoretically, and typically formulate their market segmentation strategy in pursuit of maximizing their own revenue. As such, we propose to study this issue with an optimization model.

The remainder of the paper is arranged as follows: Section 2 briefly reviews the related literature. In Section 3, we first briefly introduce our research problem, and then propose our market segmentation model and its solution algorithm. In Section 4, we propose to use a computational experiment approach to solve our proposed model, and design computational experiments to validate our model. Section 5 discusses the managerial insights of our research findings for DSPs. Section 6 concludes our efforts.

2. Literature review

Market segmentation is one of the most effective way in marketing and advertising. It allows the marketers or advertisers to divide the consumers into diverse segments, and then choose the best-matched segments to their products or services. As such, it can greatly improve the marketing and advertising efficiency and effectiveness. In literature, the market segmentation issues have been intensively studied with many market segmentation criteria proposed.

2.1. Market segmentation in marketing research

Since Smith (1956) proposed the concept of market segmentation, researchers have realized the importance of market segmentation and the research opportunities in analyzing the sales potentials among different consumer groups in marketing (Claycomb and William, 1968). Based on the concept of market segmentation, Kotler (1994) proposed the segmentation–targeting–positioning (STP) approach, which has become one of the most important theories in marketing. In practice, companies can segment their customers into multiple groups with the help of data mining techniques such as cluster analysis (Berry and Linoff, 2014), of which K-means cluster analysis is the most often used method (Wedel and Kamakura, 2012). In addition, the RFM (Recency, Frequency, Monetary) method proposed by Hughes (1994) is an effective method for market segmentation in marketing, which segment the audiences through analyzing their length of a time period since the last purchase (R), number of purchase within a specified time period (F) and amount of money spent in this specified time period (M) (Wei et al., 2010). Generally, different groups differ in their attributes, needs and behavior (Kotler et al., 2005). Thus, the most important issue in market segmentation is selecting a suitable criteria to segment users, which directly determines the segmentation strategy and marketing effects.

In literature, the criteria for segmenting the users were widely studied. Kucukmiroglu (1999) proposed a market segmentation approach using the lifestyle patterns and ethnocentrism of the con-
sumers. Blattberg and Sen (1974) proposed to take into consideration the multi-dimensional purchasing behavior of the consumers, which can divide the consumers into multiple homogeneous purchase segments, and will direct the company’s marketing efforts to the appropriate segments. Considering the brand preference and price sensitivity of the consumers, Kamakura and Russell (1989) proposed a flexible probabilistic choice model for market segmentation, which can partition the market into consumer segments with different brand preferences and price sensitivities. With massive amount of data generated in tourism markets, the tourists can be divided into groups sharing different characteristics and tailored marketing strategies (Dolnicar, 2002; Dolnicar and Bettina, 2008; Dolnicar et al., 2012; Zins, 2008).

2.2. Market segmentation in advertising

In advertising research (Lo et al., 2014; Yuan and Zeng, 2012), the targeting accuracy is a major concern of the advertisers, since it can help the advertisers reach their highly matched audiences with lower costs and better Return-On-Investments(ROI). For this purpose, many targeting approaches have been proposed, in which user segmentation is regarded as the most effective way to improve the target accuracy for the advertisers (Wedel and Kamakura, 1999; Weinstein, 2013).

User segmentation aims to divide the users into multiple distinct groups, such that the users in the same group have similar interests, while the features among groups are quite different. Such clustering methods as k-means and probabilistic-density-based mixture models are feasible approaches for market segmentation (Wedel and Kamakura, 2012). In practical applications, an easy way is to utilize a simple static property such as the age, to segment the users into multiple groups with different age intervals (Mazane, 2000), and such segmentation method is particularly effective when advertising for age-specific products (Grosset and Viscolani, 2005; Faggian and Grosset, 2013). Weaver and McCleary (1984) examined the reactions of different market segments to the travel advertisements, and obtained that age is a significant segmentation variable for travel advertisers. By combining a set of features, Myers and Tauber (2011) proposed a post hoc approach to segment the market, and Prakash (1986) proposed a framework for segmenting the women’s market with personal values and the means-end chain model. Since the Q method is an effective qualitative approach, and the R method is an effective quantitative approach, Kim and Lee (2015) combined the Q theory and the R empirical method, and proposed a Q-R hybrid methodology to segment the customers in mobile advertising businesses.

Generally, the online behavior of the users can well characterize various kinds of users (Yan et al., 2009), thus many behavior-based user segmentation approaches were proposed (Chen and Stallaert, 2014). As users’ behavior contains many implicit signals about users’ interests and content attractiveness, the analysis of the historical user activity will be a great help in building behavior-driven user segmentation (Bian et al., 2015). In general, search behavior is regarded as the most useful user activity and also the most important indicator to reveal the user’s behavior pattern. Based on these user queries, Tu and Lu (2010) proposed a topic-based user segmentation algorithm, which can well divide the users with similar query terms into the same group. Moreover, a lot of hidden semantics maybe embedded in the search behavior. Thus, Gong et al. (2013) proposed a latent semantic user segmentation approach based on Latent Dirichlet Allocation (LDA), and Wu et al. (2009) proposed a semantic user segmentation strategy by adopting a Probabilistic Latent Semantic Approach (PLSA), to mine the hidden semantics and maximize the value of search behavior. Compared to the search behavior of the users, the evaluative behavior has a more powerful ability to reflect the implicit preferences of the users. By studying the semantic overlapping between the classes of items positively evaluated by the users and the rest of classes, Saia et al. (2015) proposed an interpretable and non-trivial user segmentation approach to uncover the implicit preferences, which is quite effective in finding the desired target audiences for the advertisers. With the semantic analysis on the descriptions of the items positively rated by the users, Saia et al. (2016) obtained the reliable user preferences, which can help advertisers realize semantic behavioral targeting. Besides, the purchasing behavior of online consumers is of great importance in the electronic commerce market. As, such, Wu and Chou (2011) developed a soft clustering method to classify online customers based on their purchasing data across categories.

2.3. Market segmentation in RTB advertising

As for RTB advertising, precision marketing is the most important characteristic, and is realized through the market segmentation strategies of the DSPs. Thus, the market segmentation strategy of the DSPs is of great importance.

However, possibly due to the high dependency on real world data and the increasing privacy protection concerns from RTB companies, there are currently few studies on market segmentation problem in RTB markets. The most relevant work to this paper is to study the targeting issue through user classification with multi-dimensional user portrait (Cheng et al., 2016; Zhang et al., 2016), which can only be regarded as one key step of market segmentation.

However, our work is focused on not only the user portrait, but also on the granularity of the user portrait, as well as how the changes of the granularity affect the total revenues of the advertisers. Thus, this paper can be regarded as an important step further towards a unified framework to study the market segmentation issue, which has not been studied in the literature.

Despite of the absence of research efforts, market segmentation has attracted increasing attention from the RTB industry. For example, as one of the largest DSP platforms in China, iPinyou has proposed its patented audience profiling technology (Zhang et al., 2014), and published its white book on Digital Advertising Audience Taxonomy (DAAT), which can segment the audiences into more than 6000 niche markets according to the dimensions of audience attributes, geographical distribution, personal attention and purchase intention. Thus, it is of great importance to study the market segmentation issues in the RTB markets, in order to provide meaningful management insights and actionable operational guidance for DSPs in RTB industry.

3. Model and solution

3.1. Problem statement

For simplicity, we use “strategy” and “granularity” in the following sections to represent “market segmentation strategy” and “market segmentation granularity”, respectively.

In RTB markets, DSPs typically label the target audiences (or users) with various kinds of tags, and each tag can divide the audiences into multiple groups (e.g., Fig. 1). With the number of tags increasing, a hierarchical structure of the audiences can be obtained. For example, suppose there are $M$ tags dividing the audiences into $t_1, t_2, \ldots, t_M$ groups, respectively. Then we have a $(M+1)$-layer structure of the audiences with no tag in the top level, and one more tag from the upper level to the lower level, as shown in Fig. 2. Obviously, the granularity increases from the top level to the bottom level, and correspondingly the number of
users in each niche market decreases, resulting in better matching and targeting for advertisers.

For DSPs, each level corresponds to one of their strategies. A DSP has to choose an optimal strategy, so as to maximize the marketing effect for all advertisers.

Under each market segmentation strategy, the auction process for an ad request is illustrated in Fig. 3, which can be described as follows:

1. Under each market segmentation strategy, the DSP segments the market into multiple niche markets, and find the niche market corresponding to each advertiser.
2. As soon as an ad request arrives from the AdX, the DSP decides the corresponding niche market of the ad request. If the advertisers are located in niche market corresponding to the ad request, then the ad request is matched with the advertiser, and the advertiser will bid for the ad request.
3. The DSP finds the highest bid ($v_1$ bid by advertiser 1) and second highest bid ($v_2$ bid by advertiser 2) of the bidding advertisers, and delivers them to the AdX.
4. Denote the reserve price of the ad request on the AdX as $q$, and the highest bid of all the other DSPs as $v'$. If the highest bid $v_1$ is larger than $v'$ and $q$, i.e., $v_1 > \max(v', q)$, then advertiser 1 will win the auction, and he/she need to pay $\max(v_2, v', q)$; otherwise, advertiser 1 will lose the auction.

In the following, we provide an example to illustrate the effect of different DSPs’ strategies on the targeting effect of advertisers.

**Example 1.** Suppose there are two tags to characterize the 15 target audiences of four advertisers on one DSP as given in Table 1, and each tag can divide the market into two niche markets. Correspondingly, the DSP has three strategies, adopting 0, 1 and 2 tags, respectively, as shown in Fig. 4. For Strategy-I, the market will not be divided, and all the four advertisers participate in the same auction and bidding process. For Strategy-II, the market will be divided into two niche markets by adding the tag of “gender”, resulting in two advertisers in each niche market. Similarly, in Strategy-III, a new tag “geo” is added, resulting in four niche markets with one advertiser in each market.

Obviously, the granularity will be refined with the increased number of tags. This results in the following facts: on one hand, the competition will be reduced due to the decreased number of advertisers and target audiences in each niche market; on the other hand, the advertisers’ value-per-ad-impression will be increased due to the better precision of advertiser-audience matching. Thus, DSPs should leverage the above two aspects in choosing the optimal granularity.

As can be seen in Example 1, the targeting of the advertisers will be refined with the increasing of the granularity. However, with the refinement of the granularity, advertisers may also risk losing some potential customers. For example, suppose the customers have an age span ranging from 18 to 38, and the potential audiences of the advertiser are located in the age span $A = [18, 28]$. If the DSP segments the customers into four segments with age spans as $A_1 = [18, 23], A_2 = [23, 28], A_3 = [28, 33]$ and $A_4 = [33, 38]$, respectively, then for the advertiser, choosing $A_1$ will result in a potential loss of the customers in $A_2$, and vice versa. Such dilemma has troubled Microsoft’s ADECN platform after it segmented its markets into many extremely narrow niche markets (Bergemann and Bonatti, 2011; Levin and Milgrom, 2010). To deal with such problems, we consider the case that one advertiser can choose multiple segments under a granularity in our model, i.e., the adver-
AdX

Fig. 3. The auction process for an ad request under each granularity.

Table 1
The corresponding tags of the 15 target audiences.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>geo = China</th>
<th>geo = USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender = male</td>
<td>a1, a2, ..., a5</td>
<td>a6, a7</td>
</tr>
<tr>
<td>gender = female</td>
<td>a8, a9, a10</td>
<td>a11, a12, ..., a15</td>
</tr>
</tbody>
</table>

timator can choose both A1 and A2 in the above example. As such, our solution can effectively avoid the loss of potential audiences of the advertisers.

3.2. Notations

In this section, we list the notations in our model in Table 2.

3.3. Market segmentation model

3.3.1. Segmentation granularities and strategies

Consider the scenario that there are K competing DSPs in the RTB market, i.e., the winning advertiser in one DSP will obtain the ad impression only if he/she defeats the winning advertisers on other DSPs.

Suppose there are M alternative tags, represented by \( T = \{T_1, T_2, ..., T_M\} \), corresponding to the target audiences, and each tag \( T_i \) can segment the audience into \( t \) groups. Generally, the tags with greater importance can have better performance. The importance of the tags can be evaluated by the relevance with the characteristics of the advertisers’ target audiences and ability to discriminate these ad requests. For convenience, we assume the set \( T \) is arranged with decreasing importance. With these tags, DSPs can segment the RTB market with diverse granularities, and thus lead to different number of niche markets. For example, if the DSP chooses none tag, then there is no segmentation to the market. In this case, there is only one market; if the DSP chooses the tags \( \{T_1, T_2, ..., T_i\} \), then the market will be divided into \( \prod_{j=1}^i T_j \) niche markets. Denote all the possible granularities as \( \mathcal{L} = \{L_1, L_2, ..., L_M, 1\} \), where \( L_i \) is generated with \( i - 1 \) tags, and if we let \( t_0 = 1 \), then \( L_1 \) corresponds to \( \prod_{j=1}^1 T_j \) niche markets. Thus, in the granularity \( L_i \), there is only one market, and \( L_i (i \geq 2) \) can be represented by an \( i - 1 \) dimension function

\[
F_i(f_1, ..., f_{i-1}) = f_{i-1} + \sum_{j=1}^{i-2} (f_j - 1) \prod_{j=k+1}^i t_k
\]

where \( f_j \in \{1, 2, ..., t_i\} \). Obviously, \( F_i(\cdot) \) has \( \prod_{m<i} t_m \) values, each corresponding to a niche market under granularity \( L_i \). In practice, it is easy to obtain the corresponding niche market according to the function \( F_i(f_1, f_2, ..., f_{i-1}) \) when the values of \( f_j, j < i \) are given.

Denote \( Y_i = \{1, 2, ..., \prod_{j=1}^{i-1} t_j\} \), then \( Y_i \) represents the set of niche markets under granularity \( L_i \), and \( J \in Y_i \) represents the \( j \)-th niche market under granularity \( L_i \). With \( M \) tags, there are \( M + 1 \) granularities, corresponding to which, the DSP has \( M + 1 \) strategies to segment the market.

3.3.2. Matching ad requests and advertisers

For a certain period of time, suppose there are \( S \) ad requests, denoted as \( Q = \{q_1, q_2, ..., q_s\} \), and for each request \( q_j \in Q \), the reserve price is \( p_j \). With the \( M \) tags, each ad request can be represented by a vector \( q_j = \{q_{j1}, q_{j2}, ..., q_{jM}\} \), where \( q_{jm} \in \{1, 2, ..., t_m\} \) represents the value of \( q_j \) at tag \( T_m \). For simplicity, we assume \( q_{jm} \) takes only one value, and then each ad request falls into only one niche market under each granularity. Denote the niche market of ad request \( q_j \) under granularity \( L_i \) as \( Y_{ij} \), which is a function on \( L_i \) and \( j \), and can be computed through function \( F_i(\cdot) \) in (1) as follows

\[
y_{ij} = F_i(f_1, f_2, ..., f_{i-1}) = \sum_{j=1}^{i-2} (q_{jm} - 1) \prod_{j=k+1}^i t_k
\]

Obviously, it takes the values in the set

\[
Y_i = \{1, 2, ..., \prod_{j=1}^{i-1} t_j\}
\]

Denote all the advertisers on the DSP as \( U = \{u_1, u_2, ..., u_N\} \), and their total budgets as \( B = \{B_1, B_2, ..., B_N\} \). Under each granularity \( L_i \), the advertiser \( u_k \in U \) should choose his/her matched niche markets through the previously defined function \( F_i(\cdot) \) in (1). Suppose there are \( \beta_{ik} \) niche markets matched with the advertiser \( u_k \) under granularity \( L_i \), denote as \( D_k = \{z_{ik1}, z_{ik2}, ..., z_{ik\beta_{ik}}\} \), where \( z_{ik\eta} = 1, 2, ..., \beta_{ik} \) represent the \( \beta_{ik} \)-th niche market corresponding to advertiser \( u_k \). Obviously, \( D_k \subseteq Y_i \).

With the \( M \) tags, each advertiser \( u_k \) can be represented by a vector \( u_k = (u_{k1}, u_{k2}, ..., u_{kM}) \), where \( u_{km} \in A_{km} \subseteq \{1, 2, ..., t_m\} \) represents the value set of \( u_k \) at tag \( T_m \). Since each \( u_k \) at least takes one
value at each tag, if we denote $N_{k,x} = |A_{k,x}|$, then $N_{k,x}$ represents the number of corresponding groups of the audience for advertiser $u_k$ at tag $T_x$. Thus, if we let $N_{k,0} = 1$, then the number of niche markets corresponding to advertiser $u_k$ at each granularity $L_i$ can be computed by $b_{i,k} = \prod_{x \in T_i} N_{k,x}$. In the following, we compute each niche market of advertiser $u_k$ under granularity $L_i$ through function $F_i(\cdot)$ in (1).

Denote $A_{k,x}$ as

$$A_{k,x} = \{a_{k,x,1}, \ldots, a_{k,x,N_{k,x}}\}. \quad (3)$$

---

**Fig. 4.** An illustrating example for market segmentation with two tags and four advertisers.
Let $\sigma_i(k,j) \in (0,1)$ be the matching probability between the advertiser $u_k \in U$ and the ad request $q_j \in Q$ at granularity $L_i$. Then $\sigma_i(k,j) = 1$ if and only if advertiser $u_k$ and ad request $q_j$ fall into the same niche market under granularity $L_i$, i.e., there exists some $p_k \in \{1, 2, \ldots, \beta_{L_i}\}$ such that $y_j = z_{i,k,p}$, from which we can obtain that $y_j \in D_{i,k}$. Thus, the matching probability $\sigma_i(k,j)$ of advertiser $u_k$ and ad request $q_j$ can be computed as follows

$$
\sigma_i(k,j) = \begin{cases} 
1, & \text{if } y_j \in D_{i,k} \\
0, & \text{otherwise.}
\end{cases}
$$

Generally, the matching probability is a measurement for the matching degree of the ad request (and also the audience) with the advertiser. The advertiser $u_k$ will participate in the auction of $q_j$ under granularity $L_i$ only if $\sigma_i(k,j) = 1$.

Moreover, when $\sigma_i(k,j) = 1$, we can obtain that the value of ad request $q_j$ to advertiser $u_k$ is $v_i(k,y_j)$, and the bid of advertiser $u_k$ to ad request $q_j$ is also $v_i(k,y_j)$.

### 3.3.3. Modeling the two-stage auction

The first stage auction is run by the DSP among all the bidding advertisers. First, we need to find the set of advertisers bidding for each ad request. For the ad request $q_j$, if the advertiser $u_k$ has a matching probability 1 with ad request $q_j$ (e.g., $\sigma_i(k,j) = 1$), and his/her remaining budget is enough for the auction (e.g., $b_i(k,j - 1) \geq v_i(k,y_j)$), then he/she will participate in the bidding process of ad request $q_j$. Thus, the set of advertisers on the DSP bidding for ad request $q_j$ can be computed as follows

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

and the advertisers with the highest bid and the second highest bid can be obtained by

$$u_{k'}(j) = \text{argmax}_{u_k \in U_i(j)} v_i(k,y_j),$$

and

$$u_{k''}(j) = \text{argmax}_{u_k \in U_i(j) \setminus \{u_{k'}\}} v_i(k,y_j),$$

respectively.

The second stage auction is run by the AdX platform among all the DSPs. Due to the optimality of the Vickery mechanism (Myerson, 1981), most of the AdX platforms are now using the Vickery auction, where the highest bidder wins by truthfully bidding its value, but pays the second-highest bid (Yuan et al., 2014). However, due to the existence of DSPs in RTB advertising and the two-stage auction, there is no incentive for a DSP to truthfully submit all bids from her advertisers (especially the second-highest bid), which may lead to a decreased revenue for AdXs (Muthukrishnan, 2009). Fortunately, the OSP (Optional Second-

### Table 2

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>The set of tags corresponding to the target audiences, $T = {T_1, T_2, \ldots, T_T}$</td>
</tr>
<tr>
<td>$t_i$</td>
<td>The number of groups for the target audiences divided by the tag $T_i$</td>
</tr>
<tr>
<td>$L_k$</td>
<td>The set of all the possible granularities, $L = {L_1, L_2, \ldots, L_{\beta_{L_k}}}$</td>
</tr>
<tr>
<td>$\gamma_{i}^{k}$</td>
<td>The niche market set under granularity $L_i$, $\gamma_{i}^{k} = {1, 2, \ldots, \Pi_{j=0}^{i-1} \gamma_{j}}$</td>
</tr>
<tr>
<td>$Q$</td>
<td>The set of all the advertising requests during a certain time period, $Q = {q_1, q_2, \ldots, q_Q}$</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>The reserve price of ad request $q_i$</td>
</tr>
<tr>
<td>$U$</td>
<td>The set of all the advertisers on the DSP, $U = {u_1, u_2, \ldots, u_B}$</td>
</tr>
<tr>
<td>$B$</td>
<td>The total budget of all the advertisers on the DSP, $B = {B_1, B_2, \ldots, B_B}$</td>
</tr>
<tr>
<td>$y_i^{k}$</td>
<td>The corresponding niche market of ad request $q_j$ under granularity $L_k$</td>
</tr>
<tr>
<td>$D_{i,k}$</td>
<td>The corresponding niche market set of advertiser $u_k$ under granularity $L_k$, $D_{i,k} = {z_{i,k,1}, z_{i,k,2}, \ldots, z_{i,k,\beta_{L_k}}}$, where $z_{i,k,1} = 1, 2, \ldots, \beta_{L_k}$ represent the $i^{th}$ niche market corresponding to advertiser $u_k$</td>
</tr>
<tr>
<td>$\sigma_i(k,j)$</td>
<td>The matching probability between the advertiser $u_k$ and the ad request $q_j$, $\sigma_i(k,j) \in (0,1)$</td>
</tr>
<tr>
<td>$v_i(k,y_j)$</td>
<td>The value function of the advertiser $u_k$ for ad requests in niche market $z_{i,k}$ under granularity $L_k$, $\sigma_i(k,j) = 1$, and the bid of advertiser $u_k$ to ad request $q_j$</td>
</tr>
<tr>
<td>$c_i(k,j)$</td>
<td>The cost of advertiser $u_k$ to ad request $q_j$ under a granularity $L_k$</td>
</tr>
</tbody>
</table>

$$\mu_{1,i,1} = \left(\begin{array}{c} a_{k,1,1,1}, \ldots, a_{k,1,1,1} \\ a_{k,1,2,1}, \ldots, a_{k,1,2,1} \\ \vdots \\ a_{k,1,1,\gamma_{i}^{k}} \end{array} \right),$$

$$\mu_{1,i,2} = \left(\begin{array}{c} a_{k,1,1,1}, \ldots, a_{k,1,1,1} \\ a_{k,1,2,1}, \ldots, a_{k,1,2,1} \\ \vdots \\ a_{k,1,1,\gamma_{i}^{k}} \end{array} \right),$$

$$\mu_{1,i,i-1} = \left(\begin{array}{c} a_{k,1,1,1}, \ldots, a_{k,1,1,1} \\ a_{k,1,2,1}, \ldots, a_{k,1,2,1} \\ \vdots \\ a_{k,1,1,\gamma_{i}^{k}} \end{array} \right),$$

where $\gamma_{i}^{k}$, $a_{k,1,1,1}$, and $a_{k,1,2,1}$ denote the first $\gamma_{i}^{k}$ dimensions and $a_{k,1,1,1}$, $a_{k,1,2,1}$ represent the $i^{th}$ niche market corresponding to advertiser $u_k$.
Price) mechanism (Mansour et al., 2012), which has been practically used by Google DoubleClick system, can well tackle this problem, and the OSP mechanism can also force advertisers to truthfully report their bids. Moreover, as estimated by Datanyze, the market share of Google DoubleClick is 67.2% in 2017, which illustrates that at least 67.2% of the world’s ad traffic are sold using the OSP mechanism. Thus, in this paper, we adopt the OSP mechanism in the second stage auction on the AdX.

Denote the highest bid for the ad request \( q \) on all the other DSPs as \( v'(j) \). Since the highest bid on the DSP is \( v_i(k'(i,j),y_{ij}) \), if \( v_i(k'(i,j),y_{ij}) > \max\{v'(j),\rho_j\} \), advertiser \( u_{k'(i,j)} \) on the DSP will win the auction, and otherwise, advertiser \( u_{k'(i,j)} \) will lose the auction. Thus, if we define

\[
I_i(k'(i,j),j) = \begin{cases} 1, & \text{if } v_i(k'(i,j),y_{ij}) > \max\{v'(j),\rho_j\} \\ 0, & \text{otherwise} \end{cases}
\]  

(10)

then \( I_i(k'(i,j),j) \) is the indicator function representing advertiser \( u_{k'(i,j)} \) on the DSP wins (or not) ad request \( q \) under granularity \( L_i \), and only if \( I_i(k'(i,j),j) = 1 \), advertiser \( u_{k'(i,j)} \) will win ad request \( q \).

According to the OSP mechanism, a DSP needs to submit both the top two bids of her advertisers, thus the advertiser with the highest bid wins the auction, and he/she needs to pay only the second highest bid on the AdX (Mansour et al., 2012). If advertiser \( u_{k'(i,j)} \) wins, i.e., \( I_i(k'(i,j),j) = 1 \), then he/she needs to pay \( \max\{v_i(k'(i,j),y_{ij}), v'(j), \rho_j\} \) for ad impression \( q \). Otherwise, if \( I_i(k'(i,j),j) = 0 \), advertiser \( u_{k'(i,j)} \) can not win ad impression \( q \), and his/her cost is 0. Thus, we can define the cost function of advertiser \( u_{k'(i,j)} \) as

\[
c_i(k'(i,j),j) = \max\{v_i(k'(i,j),y_{ij}), v'(j), \rho_j\} I_i(k'(i,j),j),
\]  

(11)

and the remaining budget for advertiser \( u_i \) after ad request \( q \) is

\[
b_i(k,j) = \begin{cases} b_i(k,j-1) - c_i(k'(i,j),j), & \text{if } k = k'(i,j) \\ b_i(k,j-1), & \text{otherwise} \end{cases}
\]  

(12)

where \( b_i(k,0) = B_k \).

3.3.4 Market segmentation model

Assume the revenue of the advertiser from an ad request is equal to the advertiser’s value, then if \( I_i(k'(i,j),j) = 1 \), the advertiser \( u_{k'(i,j)} \) will win ad impression \( q \), and obtain \( v_i(k'(i,j),y_{ij}) \) revenue, and otherwise, his/her revenue will be 0. Denote the revenue of all advertisers on the DSP from ad request \( q \) under the granularity \( L_i \) as \( r_i(j) \), then we have

\[
r_i(j) = v_i(k'(i,j),y_{ij}) I_i(k'(i,j),j).
\]  

(13)

Thus, under granularity \( L_i \), the total revenue of the advertisers on the DSP from all the ad requests in \( Q \) can be computed by

\[
g(i) = \sum_{j=0}^{M} r_i(j).
\]  

(14)

The DSP aims to choose the optimal granularity from \( L_i \) to maximize the total revenue of all the advertisers, i.e.,

\[
\max_{L_i} g(i).
\]  

(15)

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

\[
\begin{align*}
\max_{L_i} & g(i) = \sum_{j=0}^{M} r_i(k'(i,j),j)I_i(k'(i,j),j) \\
\text{subject to :} & \\
D_{ik} & = \{z_{ik1}, z_{ik2}, \ldots, z_{ikA_k}\} \\
\sigma_i(k,j) & = \begin{cases} 1, & \text{if } y_{ij} \in D_{ik} \\ 0, & \text{otherwise} \end{cases} \\
U_i(j) & = \{u_i | u_i \in U, \sigma_i(k,j) = 1, b_i(k,j) > v_i(k,y_{ij})\} \\
u_{ik}(j) & = \arg\max_{u_i \in U_i(j)} \{v_i(k,y_{ij})\} \\
u_{ik'}(j) & = \arg\max_{u_i' \in U_i(j)} \{v_i'(k',y_{ij})\} \\
I_i(k'(i,j),j) & = \begin{cases} 1, & \text{if } v_i(k'(i,j),y_{ij}) > \max\{v'(j),\rho_j\} \\ 0, & \text{otherwise} \end{cases} \\
c_i(k'(i,j),j) & = \max\{v_i(k'(i,j),y_{ij}), v'(j), \rho_j\} I_i(k'(i,j),j) \\
b_i(k,j) & = \begin{cases} b_i(k,j-1) - c_i(k'(i,j),j), & \text{if } k = k'(i,j) \\ b_i(k,j-1), & \text{otherwise} \end{cases} \\
b_i(k,0) & = B_k
\end{align*}
\]  

(16)

where \( y_{ij} \) and \( z_{ikA} \) are given in (2) and (5), respectively, \( D_{ik} \) represents the set of the corresponding niche market of advertiser \( u_i \) at granularity \( L_i \), \( \sigma_i(k,j) \) represents the matching probability of advertiser \( u_i \) and ad request \( q_i \) at granularity \( L_i \), \( u_{ik'(i,j)} \) and \( u_{ik'}(j) \) represent the advertisers with the highest bid and the second highest bid for ad request \( q_i \) among the advertisers on the DSP at granularity \( L_i \), \( I_i(k'(i,j),j) \) is the indicator function representing whether the winning advertiser \( u_{ik'(i,j)} \) on the DSP win or not ad request \( q_i \) at granularity \( L_i \), \( c_i(k'(i,j),j) \) represents the cost of advertiser \( u_{ik'(i,j)} \) for ad request \( q_i \) at granularity \( L_i \), \( b_i(k,j) \) represents the remaining budget of advertiser \( u_i \) after ad request \( q_i \) at granularity \( L_i \), and \( b_i(k,0) \) represents the initial budget of advertiser \( u_i \) at granularity \( L_i \).

Solving the above model, we can obtain the optimal granularity \( L_i \), and the corresponding optimal revenue is \( g(i) \).

3.4 Solution algorithm

The proposed model (16) is difficult to solve theoretically, since it contains a lot of complex iterative processes for each granularity and each ad request. Moreover, the iterative processes for each ad request is highly depend on the result of the previous ad request. Thus, the numerical solution of model (16) can hardly be obtained exactly.

3.4.1 Description of the algorithm

In order to find the optimal granularity, we should first compute the revenue of the advertisers from each ad request at each granularity, then the total revenue in each granularity can be obtained by adding up the revenues from all the ad requests. Finally, we can obtain the maximized revenue and the corresponding optimal granularity by comparing the total revenues at all the granularities. Thus, the detailed solving process of model (16) can be described as follows:

**Step 1:** Under each granularity \( L_i \), compute the matching probability \( \sigma_i(j,k) \) for each ad request \( q_i \) with each advertiser \( u_i \) according to formula (6), and find the set of advertisers \( U_i(j) \) on the DSP bidding for ad request \( q_i \) according to formula (7).

**Step 2:** Find the advertisers with the highest and the second highest bids \( u_{ik'(i,j)} \) and \( u_{ik'}(j) \) on the DSP for ad request \( q_i \) from the advertiser set \( U_i(j) \) according to formulas (8) and (9).
Step 3: Check if advertiser $u_{i,j}$ wins ad request $q_i$ or not by comparing his/her bid with the highest bid of all the winning advertisers on other DSPs, according to formula (10).

Step 4: Compute the cost $c_i(k(i,j),j)$ of advertiser $u_{i,j}$ for ad request $q_i$, according to formula (11), and update the remaining budget $b_i(k,j)$ of each advertiser $u_i$ according to formula (12).

Step 5: Compute the revenue on the DSP from ad request $q_i$ under granularity $L_t$ according to formula (13).

Step 6: Get the total revenue $g(i)$ on the DSP by adding up the revenues from all the ad requests $q_i \in Q$ according to formula (14).

Step 7: Find the maximized total revenue $g(i')$ and the corresponding optimal granularity $L_{t'}$ according to formula (15).

Based on the above discussions, we propose a solution algorithm to find the optimal granularity $L_{t'}$, as given in Algorithm 1.

**Algorithm 1 An algorithm for granularity optimization**

1. **Input Data**: $L, Q, U, Y_i, D_{i,k}, y_{i,j}
2. **Output Result**: $L_{t'}, V
3. **initialization**: $i = 1, i' = 1, V' = 0;
4. **repeat**
5. 
6. **for** $k = 1, N$ **do**
7. 
8. **end for**
9. **for** $j = 1, S$ **do**
10. 
11. **for** $k = 1, N$ **do**
12. 
13. **end if**
14. 
15. **end for**
16. 
17. 
18. 
19. 
20. **else** $l'(k(i,j),j) = 0$;
21. 
22. **end if**
23. 
24. **if** $k = k'(i,j)$ **then**
25. 
26. **else**
27. 
28. **end if**
29. 
30. **end for**
31. 
32. 
33. **end if**
34. 
35. **until** $i = M + 2$

3.4.2. Optimality of the algorithm

In Algorithm 1, the time complexity is $O(M \times S + N)$ since there are 3-layer nested cycle, and the order of magnitudes of their cycle numbers are $M, S, N$, respectively. Moreover, only when seeking for $u_{i,j}$ and $u_{i,j}$ in $U_i(j)$ in each cycle, one unit of additional memory space is needed in the sorting process. Thus, the space complexity of Algorithm 1 is $O(1)$. Therefore, Algorithm 1 can be regarded as an efficient algorithm with polynomial complexity.

Obviously, Algorithm 1 is a heuristic algorithm, and in the following, we discuss its optimality. As can be seen from model (16), when seeking for the optimal granularity, we need to know the values of the following parameters:

1. Each granularity $L_t$: the set of tags $\{T_1, T_2, \ldots, T_{t-1}\}$ used in granularity $L_t$, all the niche markets $Y_t = \{1, 2, \ldots, \prod_{i=1}^{t-1} f(i)\}$ under $L_t$;
2. All the advertisers on the DSP: the set of advertisers $U = \{u_1, u_2, \ldots, u_k\}$, the corresponding total budget $B = \{B_1, B_2, \ldots, B_k\}$, the representations of each advertiser $k$ with these tags $u_k = (u_{k,1}, u_{k,2}, \ldots, u_{k,t-1})$;
3. The sequence of the coming ad impressions: The set of the ad impressions $Q = \{q_1, q_2, \ldots, q_k\}$, the reserve price $\rho_j$ for each request $q_j$, the representations of each ad impression $q_j$ with these tags $q_{j} = (q_{j,1}, q_{j,2}, \ldots, q_{j,t-1})$;
4. The highest bids of other DSPs: For each ad impression $q_j$, the highest bids $v'(j)$ of all the other DSPs.

When the values of the above parameters are given, the results of the two-stage auction under each granularity are actually determined, and the total revenues of the advertisers can be exactly calculated. Since there are $M$ tags, the segmentation granularity has $M + 1$ feasible values. Thus, we can obtain the exact value of the total revenues of the advertisers under each of the $M + 1$ granularities by carrying out the above process $M + 1$ times. Therefore, model (16) can be regarded as an integer programming with the feasible solution set as $\{1, 2, \ldots, M + 1\}$, and Algorithm 1 is an exhaustive algorithm for solving the integer programming. Due to the finiteness of $M + 1$, Algorithm 1 is an effective solution algorithm for finding the optimal solution of model (16).

4. Computational experiments

The computational experiment approach is first successfully used in solving economic issues by Kydland and Prescott (1996) in 1996. After the innovative ACP (Artificial Societies + Computational Experiments + Parallel Systems) theory proposed by Wang (2004), computational experiments have been recognized by researchers and widely used in dealing with various kinds of socio-economic problems with success. Due to the essential complexity of online RTB markets, it is difficult to validate our proposed model and strategies with online field experiments. Fortunately, computational experiments can serve as an alternative way and has been successfully applied in complex systems. Thus, in this section, we utilize the computational experiment approach to validate our proposed model.

In our experiments, we aim to validate the effect of market segmentation on the advertising effect of the advertisers on the DSP, and explore the influence factors of the optimal market segmentation strategy. For this purpose, we design the following two experiments.

4.1. Finding the optimal granularity

4.1.1. Computational experiment scenario

Suppose there are 2 homogeneous DSPs in the market adopting the same market segmentation strategy, and thus the winning probabilities of the two DSPs can be assumed to be the same. There are altogether 13 tags to characterize each ad request, and each tag can divide the ad requests into 2 groups. With these tags, the DSP
has 14 feasible strategies, with the number of tags indexed as 0 to 13, as shown in Fig. 5. The numbers of niche markets under these strategies are $2^0 = 1, 2^1 = 2, 2^2 = 4, \ldots, 2^{13} = 8192$, respectively. The purpose of the DSP is to find the optimal granularity to maximize the total revenues of the advertisers.

In practice, the advertisers may have one or multiple, fixed or varied target niche markets under each granularity. Thus, without lose of generality, we design three computational experimental scenarios with the computational experiment approach, which can be described as follows:

1. **Scenario-One**: Each advertiser has only one target niche market under each strategy.
2. **Scenario-Fixed**: Each advertiser can have multiple target niche markets under each strategy, but the number of niche markets of all the advertisers is the same and fixed under Strategy-14.
3. **Scenario-Random**: A more general case that each advertiser can have multiple target niche markets under each strategy, but the numbers of niche markets of the advertisers can be different, which are randomly generated under Strategy-14.

4.1.2. Computational experimental data

To evaluate the 14 strategies, we randomly generate an experimental scenario that contains 2 DSPs, 2,000,000 ad requests and 100 advertisers. These ad requests are randomly distributed in these niche markets, and the total budgets of the advertisers are uniformly distributed in $[200, 1500]$. Fig. 6 randomly generates the above data in our experiment.

For Scenario-One, the target niche market for each advertiser under Strategy-14 is generated randomly from these niche markets, as shown in Fig. 7. For Scenario-Fixed, without loss of generality, we can set the number of the corresponding niche markets for each advertiser under Strategy-14 as 2, and the randomly generated corresponding niche markets of each advertiser under Strategy-14 are given in Fig. 8. For Scenario-Random, we randomly generate the number of the corresponding niche markets for each advertiser under Strategy-14, which are assumed to take the values from 1 to 3, and then the corresponding niche markets of each advertiser under Strategy-14 can be randomly generated, as shown in Fig. 9.

Since advertisers’ values of ad impressions typically increase with the accuracy of matching (Mobius et al., 2012), we can assume that the CPMs of the advertisers increase with the granularity. Suppose the CPMs of each advertiser under Strategy-1 is uniformly distributed in $[2.00, 5.00]$, and with the increasing of the granularity, the lower bound and the upper bound of $z$ will increase 1.00 and 2.00 each time, respectively. Moreover, we assume that the advertiser with a higher CPM under one strategy also has a higher CPM under other strategies.

4.1.3. Experimental result

In order to obtain general conclusions for the optimal strategy, we conduct 2000 independent experiments for Scenario-One, Scenario-Fixed and Scenario-Random, respectively, and the total revenues of the advertisers on the DSP in these 2000 experiments are shown in Figs. 10–12, respectively. Moreover, with the standard deviation of the total revenues, we can obtain their error
Fig. 7. The target niche market of each advertiser under Strategy-14 for Scenario-One.

Fig. 8. The target niche markets of each advertiser under Strategy-14 for Scenario-Fixed.

Fig. 9. The target niche markets of each advertiser under Strategy-14.

Fig. 10. The total revenue of the advertisers on the DSP in 2000 experiments for Scenario-One.

Fig. 11. The total revenue of the advertisers on the DSP in 2000 experiments for Scenario-Fixed.

Fig. 12. The total revenue of the advertisers on the DSP in 2000 experiments for Scenario-Random.
In all the 2000 experiments for each scenario, the total revenues have a tendency of a rise first and followed by a decline, with the increasing of the granularity. The maximum occurs at Strategy-6 1996 times (99.8%) for Scenario-One, Strategy-6 76 times (3.8%) and Strategy-7 1924 times (96.2%) for Scenario-Fixed, and Strategy-6 595 times (29.7%) and Strategy-7 1405 times (70.25%) for Scenario-Random. This indicates that there exists a threshold (6 for Scenario-One, 7 for Scenario-Fixed and Scenario-Random); when the granularity is less than the threshold, the revenues can be improved with the increasing of the granularity. However, when the granularity exceeds the threshold, the revenues will decrease sharply with the increasing of the granularity.

(2) The average total revenue for the advertisers also has a tendency of a rise first followed by a decline, with the increasing of the granularity, and the maximum occurs at the granularity under Strategy-6 for Scenario-One, and Strategy-7 for Scenario-Fixed and Scenario-Random. It is obvious that Strategy-6 is the optimal strategy for the DSP in the case of Scenario-One, and Strategy-7 is the optimal strategy for Scenario-Fixed and Scenario-Random.

(3) The variation tendency of the standard deviations of the total revenues is the same with that of the average total revenue, and the maximum is also reached at the optimal strategy for each scenario, which indicates that the total revenues have larger fluctuations at the optimal strategy than other strategies.

4.2. Exploring influence factors of optimal granularity

In the following, we aim to find the effect of the number of advertisers \( N \) as well as the number of ad requests \( S \) on the optimal strategy. The experimental scenario and the experimental data setting of this experiment are the same with those in Section 4.1, expect the number of advertisers in Section 4.2.1 and the number of ad requests in Section 4.2.2.

4.2.1. Number of advertisers

Keeping the number of ad requests \( S \) fixed as 2,000,000, we change the number of advertisers \( N \) from 1 to 2000, and the effect of the number of advertisers on the total revenue under different scenarios is given in Fig. 14–16, from which we can obtain the following results:

(1) For all the three scenarios, when the number of ad requests is fixed, the average total revenue of the advertisers on the DSP will increase with the increasing of the number of advertisers under each strategy.

(2) For all the three scenarios, the optimal granularity can be affected by the number of advertisers, and a larger number of advertisers will result in a larger optimal granularity.

4.2.2. Number of ad requests

Keeping the number of advertisers \( N \) fixed as 100, we change the number of ad requests \( S \) from 1,000,000 to 12,000,000, and the effect of the number of ad requests on the total revenue under different scenarios is given in Fig. 17–19, from which we can obtain the following results:

(1) For all the three scenarios, when the number of advertisers is fixed, the average total revenue will increase with the increasing of the number of ad requests under each strategy.

(2) For all the three scenarios, with the increasing of the number of ad requests, the optimal segmentation strategy almost keeps the same, thus, the number of ad requests has little effect on the optimal granularity.

4.2.3. Result analysis

From the above experiments, we can obtain the following conclusions:

(1) For all the three scenarios, both increasing the number of advertisers on the DSP and the number of ad requests can increase the total revenue of the advertisers.

(2) For all the three scenarios, the number of advertisers has a significant effect on the optimal segmentation strategy, while the number of ad requests has little effect on the optimal granularity. Thus, the DSP should pay more attention to the number of advertisers and adjust the segmentation strategy accordingly when the number of advertisers changes.

5. Managerial insights and discussions

5.1. Managerial insights

Our research can offer useful managerial insights for DSPs in RTB markets. First, big-data driven user targeting is the most important basis for the precision marketing in RTB advertising. For DSPs and advertisers, bigger data are usually expected to bring higher revenues because of the higher matching quality between the advertisers and their audiences. However, our work indicates that this is not always true in RTB advertising auctions, due to the reduced competition among the advertisers. In our work, we validated the existence of the optimal granularity for RTB market segmentation, and once the granularity is larger than the optimal granularity, the revenues will decrease greatly. Thus, it is extremely important for DSPs to recognize that bigger data is not always better in RTB advertising, and the maximized revenues will drift away if the online big data was overused. This can explain the empirical study that after Microsoft’s ADECN platform segments its market into large numbers of niches with only 2–3 advertisers, the advertising performance is witnessed to drop significantly due to the reduced competition (Bergemann and Bonatti, 2011; Levin and Milgrom, 2010). In addition, it is widely considered that big data always means big value in practice, however, our work indi-
Fig. 14. The effect of the number of advertisers on the optimal strategy for Scenario-One.

Fig. 15. The effect of the number of advertisers on the optimal strategy for Scenario-Fixed.

Fig. 16. The effect of the number of advertisers on the optimal strategy for Scenario-Random.
Fig. 17. The effect of the number of ad requests on the optimal strategy for Scenario-One.

Fig. 18. The effect of the number of ad requests on the optimal strategy for Scenario-Fixed.

Fig. 19. The effect of the number of ad requests on the optimal strategy for Scenario-Random.
cates that the big value of big data can be realized only when they are used rationally.

Second, our work indicates that simply taking the total revenues of all the advertisers as the optimization objective is appropriate, effective and easy to realize for the DSPs, especially for such DSPs that earning their revenues by charging a commission from the advertisers, since in such cases higher revenues of the advertisers also mean higher revenues of the DSPs. Our work indicates that both increasing the number of advertisers and increasing the number of ad requests can greatly improve the total revenues of the advertisers. Thus, two possible ways can be adopted by the DSPs to improve their revenues, i.e., the one is to constantly optimize their advertising targeting effect so as to attract more advertisers, and the other is to gradually improve their capacity and disposing ability of ad traffic to switch in more ad requests. In practice, the two ways can easily be realized through market segmentation strategies. On one hand, better market segmentation strategy can improve the advertising effect for advertisers, and thus can attract more advertisers. On the other hand, better market segmentation strategies can improve the efficiency in dealing with the ad impressions, and thus with the optimization of the market segmentation strategies, the ability of disposing ad impressions can be strengthened. As such, our research results suggest that DSPs should increase the number of their advertisers and ad impressions in order to improve their revenues. The above result can well explain that why the DSP platforms iPinyou has paid special attention to develop its patented audience profiling technology in recent years (Zhang et al., 2014), continuously improved its white book on digital advertising audience taxonomy, and provided 15 updates in just 4 years from 2012 to 2015.

Third, due to the dynamics of RTB markets, the optimality of the market segmentation strategy can be changed over different periods. In such cases, it is urgent for DSPs to explore the key factors that can influence the optimality. This work indicates that the number of advertisers on the DSP can greatly influence the optimality of the market segmentation decisions of the DSP, while the number of ad requests has little influence on the optimality of granularity. Thus DSPs should pay more attention to the number of advertisers rather than the number of ad requests in making their market segmentation decisions, and adjust their strategies when the number of advertisers is greatly changed. Specifically, in case when there are only a few advertisers, over-refined granularities are not necessary for the DSPs; while in case when there are a large number of advertisers, the DSPs should better segment the market into more niche markets. Our analysis, algorithm and conclusions in this paper can provide actionable suggestions for DSPs on when and how to adjust their market segmentation strategies, which is easy to realize since the number of advertisers can be obtained by DSPs precisely. On one hand, with the logs of the previous promotions, the DSP can get the data of its advertisers such as the set of the highest bids and second highest bids among the advertisers, the total budget of each advertiser, the number of the advertisers, the total budget of each advertiser, the number of ad requests, and thus with the optimization of the relevant parameters in our model, and then the optimal strategy through our proposed model and algorithm.

5.2. Discussions

In practice, our proposed model can be easily applied by DSPs in real RTB advertising markets, from the following two aspects: Firstly, to determine the optimal market segmentation granularity. With the big data of the audiences, DSPs can have a full knowledge of their feasible market segmentation strategies. With the needs of the advertisers, the DSP can easily decide the corresponding niche markets for each advertiser under each segmentation granularity. Once an ad request comes, the DSP can easily determine its corresponding niche market under each segmentation granularity, and then find the corresponding bidding advertisers. Thus, with the logs of previous promotions, the DSP can easily estimate the information of the relevant parameters in our model, and then the optimal market segmentation strategy can be obtained using our proposed model and algorithm.

Secondly, to adjust the market segmentation granularity when the parameters are greatly changed. The promotion period can be divided into multiple time slots, the parameters of the next time slot may be changed greatly comparing with the last time slot. Thus, the DSP can utilize our model and algorithm to find the optimal granularity of the next time slot with the parameters at the end of the last time slot, to realize the dynamic adjustment of the segmentation granularity. It should be noticed that, although our results are obtained through the computational experiment approach, they can be directly used in real RTB advertising markets, since there is usually no significant difference in the work produced by data scientists who used synthetic data as opposed to real data (Patki et al., 2016).

6. Conclusions and future works

Market segmentation is an important strategic problem for DSPs, and also an important guarantee for the effectiveness and efficiency of the emerging RTB advertising paradigm. In this paper, we characterized the target audiences with multiple tags, and established a model for optimizing the market segmentation granularity. This model can be used to help DSPs to determine the optimal granularity. We proposed to use the computational experiment approach to evaluate our model, and the experimental results show that: 1) the increasing of market segmentation degree in a certain extent can not only improve the advertising effect for DSPs, but also is good for all the advertisers; 2) there is a market segmentation cap, and the advertising effect will keep rising before the cap, but quickly drop off after the cap. 3) there exists an optimal granularity, which can be influenced by the number of advertisers on the DSP, but less by the number of ad requests.

There are several limitations in the current research that will be addressed in our future work. First, we adopted the total revenue of all the advertisers on the DSP as the optimization objective of the DSP in our proposed market segmentation model, and did not consider the revenues of DSPs and the principle-agent games between DSPs and the advertisers. However, such limitations will not undermine our contributions in this research. In RTB practice, many DSPs realize their profits by charging advertisers with a certain proportion of their revenues as commission, and in this case, higher total revenues of the advertisers will lead to higher revenues for the DSPs. As such, our model and proposed strategies can offer meaningful insights for DSPs. In our future work, we plan to extend this paper by analyzing the strategic behavior and the resulting equilibrium of the principal-agent games between the
advertisers and the DSP. Second, we will try to model this market segmentation problem as a non-cooperative game under incomplete information settings among multiple competitive DSPs, and explore the DSPs’ Nash equilibrium strategies.

Acknowledgments

We gratefully acknowledge funding support from the National Natural Science Foundation of China (Grant 71472714, 61533019, 71232006, 61233001) and the Early Career Development Award of SKLMCCS (Y6S9011F4G, Y6S9011F4H).

References


Weaver, P.A., McCleary, K.W. 1984. A market segmentation study to determine the appropriate ad/model format for travel advertising. J. Travel Res. 23 (1), 12–16.


