Front. Comput. Sci., 2019, 13(2): 333–342
https://doi.org/10.1007/s11704-017-6102-0

RESEARCH ARTICLE

Optimal bundles for sponsored search auctions via bracketing scheme

Zheng-Dong XIA, Tian-Ming BU (✉), Wen-Hui GONG

Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai 200233, China

Abstract Sponsored search auction has been recently studied and auctioneer's revenue is an important consideration in probabilistic single-item second-price auctions. Some papers have analyzed the revenue maximization problem on different methods to bundle contexts. In this paper, we propose a more flexible and natural method which is called the bracketing method. We prove that finding a bracketing scheme that maximizes the auctioneer's revenue is strongly NP-hard. Then, a heuristic algorithm is given. Experiments on three test cases show that the revenue of the optimal bracketing scheme is very close to the optimal revenue without any bundling constraint, and the heuristic algorithm performs very well. Finally, we consider a simpler model that for each row in the valuation matrix, the non-zero cells have the same value. We prove that the revenue maximization problem with K-anonymous signaling scheme and cardinality constrained signaling scheme in this simpler model are both NP-hard.

Keywords sponsored search auction, revenue maximization, bracketing scheme, NP-hardness

1 Introduction

Internet advertising is a new form of advertising. The main advantages include more variants, more audience, and cheaper price. Among all variants, sponsored search is the most important. Compared with others, sponsored search can target users more easily through their inputted keywords which can be used to understand their search intentions.

The market size of sponsored search in China is 857.5 billion yuan in 2014, which is increased by 32% and accounts for 54.5% of the whole of Internet advertising market in China. More than half of the revenue of search engine companies, such as Google and Baidu, comes from sponsored search. The revenue of sponsored search in America accounts for 43% of total advertisements' revenue in 2013 [1].

There are three roles in sponsored search:

- Search engine: it is also called "auctioneer" or "matchmaker". It is responsible for the whole auction process of sponsored search. After a user submits a query to a search engine, the search engine will display the SERP (Search Engine Result Page) that contains search results and advertisements.
- User: the agent that submits the query.
- Advertiser: it is also called "bidder". When a user submits a query, advertisers bid for displaying their ads. The winner displays his advertisement in the advertisement slot.

A advertiser need submit advertisement information (such as title, description, url) to the search engine in advance, tell the search engine keywords that he is interested in, and give the bid values for them. There are some tools such as Google Adwords and Baidu Wangmeng to help advertisers to submit ads easily. The winner pays the auctioneer only when the user clicks his ads.

In general, the auction mechanism of sponsored search is GSP (Generalized Second Price Auction). GSP is a non-truthful auction mechanism for multiple items, which is an

Received February 18, 2016; accepted March 24, 2017

E-mail: tmbu@sei.ecnu.edu.cn

extension of Vickery auction. In GSP the highest n bidders win n ads slots. The ith highest bidder pays the bid value of the $(i+1)$th highest bidder. For example, there are three bidders A, B, C and two slots. If A bids five yuan, B bids two yuan, C bids one yuan, then A wins the first slot, B wins the second slot. And A pays two yuan which is the bid value of B, and B pays one yuan which is the bid value of C.

The goals of sponsored search are to help users to find their wanted information, to help advertisers to target appropriate users, and to help the auctioneer to earn huge revenue. However, these goals cannot be satisfied simultaneously. Therefore, there is a trade off among them. In this paper, we will propose a more feasible and user-friendly model which not only improves the user's experience, but also optimizes the auctioneer's revenue. More introductions of sponsored search can be found in [2, 3]. The latest research progress of sponsored search is surveyed by Qin et. al. [4].

• **Motivations** Bidding only relying on keywords may not target the user precisely enough. For example, if the keyword is "apple", then advertisers cannot determine whether it refers to the corporation or the fruit. So it will disturb advertisers to target right users. However, query is often accompanied with context [5]. A context is the demographic information of a user such as location and job. In the previous example, if advertisers know the user is a farmer, then he can infer that the query refers to the fruit. Thus the context can help advertisers to target the visitors more accurately.

However, if every bidder can bid for all context, it may result in thin market which will reduce the auctioneer's revenue. For example, suppose there are two bidders b_1, b_2. They bid for the keyword "apple". There are two contexts "farmer" and "manager". There is only one slot for bidding. Suppose the bidding matrix is as follows:

	Farmer	Manager
b_1	1	0
b_2	0	1

In the matrix above, b_1 bids 1 for "farmer" and 0 for "manager". b_2 bids 0 for "farmer" and 1 for "manager". So if the context is "farmer", b_1 wins the slot and pays 0. If the context is "manager", b_2 wins the slot and pays 0. Whatever the context is, the auctioneer's revenue is always 0.

To solve the problem of thin market, one approach is that every bidder can only bid for bundles which are subsets of contexts, not for every context. Although the solution may decrease the precision of targeting, it could increase the auctioneer's revenue.

There have been several methods to bundle the contexts. In this paper, we will present a more flexible and natural method. Suppose there are three attributes:

- Gender: male, female;
- Job: teacher, student, engineer;
- Location: Shanghai, Beijing, Shenzhen.

Therefore, there are totally 18 contexts such as (male, student, Shanghai) and (female, engineer, Beijing). Clearly, it is not convenient and necessary for the advertisers to bid for every context. Our bundling method can be illustrated by the following graphic-user-interface designed for advertisers.

In Fig. 1, an advertiser need check boxes that he is interested in. Then the bundle is the Cartesian product of the attributes each of which consists exclusively of the values interest the advertiser. In Figure 1, the advertiser checked the boxes: "Male" in Gender attribute, "China, US" in Location attribute and "Farmer, Musician" in Job attribute. So the bundle consists of (Male, China, Farmer), (Male, China, Musician), (Male, US, Farmer), (Male, US, Musician). We call this method the bracketing method. The key advantage is that the method allows an advertiser easily chooses his favorite contexts without sacrificing the flexibility.

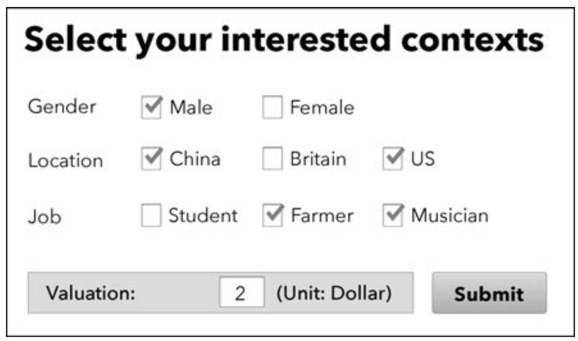


Fig. 1 GUI that needs to be filled out by advertisers

Bidding for every context will result in not only the thin market, but also the complicated process. You may imagine what a disaster it is to let an advertiser to set the bidding values of contexts one by one! So in Section 6, we will discuss a simplified model that every advertiser can only bid a fixed value for his favorite contexts.

• **Contributions** We study the revenue maximization problem under second-price mechanism in sponsored auction. Our contributions include:

1) We generalize the attribute hiding method in [13], and propose a more flexible and natural method to bundle contexts: the bracketing method. We prove that it

is strongly NP-hard to compute the optimal bracketing scheme.

2) We give a heuristic algorithm to handle the problem. By experiments on three datasets, it is shown that the heuristic algorithm performs well. Experiments also show that the optimal revenue with bracketing scheme is close to the optimal revenue without any bundling constraint.

3) On account of the complication of setting the valuation for each context one by one, we consider the constraint that for each row of the valuation matrix, the non-zero cells have the same value. We study the K-anonymous scheme and cardinality constrained scheme with this constraint and prove both of they are NP-hard to compute the optimal revenue.

• **Related work** Even-Dar et al. [5] proposed for the first time that advertisers could bid not only for keywords but also for keywords with specific contexts. They did experiments on real data and concluded that the auctioneer's revenue, social welfare and advertiser profit under context-based bidding are increased significantly compared with keyword-based bidding. The method that every context is auctioned respectively, is called the context scheme.

Then the revenue maximization problem via bundling is studied. Ghosh et al. [6] proved that it is strongly NP-hard to compute the optimal bundling, and then gave a 1/2-approximation algorithm. From another perspective, Emek et al. [7] studied the same problem. However, they called bundling a signaling scheme. Besides, they discussed a special case that the valuation matrix is a 0/1 matrix. They gave a polynomial revenue-maximum algorithm for this case. Miltersen et al. [8] extended the concept of pure signaling scheme in [7] to mixed signaling scheme. They proved that there is a polynomial revenue-maximum algorithm to compute the optimal mixed signaling scheme by linear programming, though [7] proved it is NP-hard to computer the optimal pure signaling scheme.

Further, the variants for different circumstances are discussed. Dughmi et al. [9] considered the communication length for each context is limited and proposed the concept of constrained signaling scheme. They studied the revenue maximization problem and social welfare maximization problem under the constrained signaling scheme with known/unknown valuation. Besides, the same authors [10] proved the hardness to approximate better than an $(e-1)/e$ factor for social welfare maximization under the communication-constrained signaling scheme with known valuation. Dughmi [11] studied the problem in Bayesian zero-sum games and proved that it is NP-hard to compute the signaling scheme to maximize a player's payoff in explicit zero sum games. Chen et al. [12] proposed a concept of K-anonymous signaling scheme which means each signal/bundle's size is at least K. They proved the hardness of the revenue maximization problem and the social welfare maximization problem under the K-anonymous signaling scheme.

Our paper is mainly inspired by Guo and Deligkas [13]. They pointed out the disadvantages of bidding for each context and introduced a bundling method which is called the attribute hiding method. In the attribute hiding method, an advertiser can choose for each attribute just one value or all values. An attribute hiding bundle is the Cartesian product of the attributes each of which has only one selected value or all values. They focused on the revenue maximization problem and proved that it is NP-hard to compute the optimal attribute hiding scheme. Compared with the attribute hiding method, our bracketing method is a generalization that an advertiser can choose for each attribute any number of values. Clearly, our method is more flexible. Thus the optimal revenue via bracketing scheme will be never worse than the optimal revenue via attribute hiding scheme. Actually, experiments show that the optimal revenue via bracketing scheme is very close to the optimal revenue without any bundling constraint.

Recently, Papadimitriou et al. [14] and Dasdan et al. [15] modeled the search engine that the input is keyword and the output is SERP. They proposed a new bidding form which is called output bidding. It is differed from bidding for the keyword or context on account of the output of search engine having more information to help advertisers to target users.

• **Organization of the paper** In Section 2, we formalize the bracketing method and the bracketing scheme. Section 3 proves that it is strongly NP-hard to compute the optimal bracketing scheme. Then we present a heuristic algorithm in Section 4. In Section 5, plenty of experiments are done to show that the heuristic algorithm performs well. Section 6 studies two special cases of the problem. Lastly, we summarize and conclude in Section 7.

2 Model

In this paper, we suppose there is only one slot to display the advertisement and n bidders (advertisers) to bid for the slot. There are m attributes. Each attribute $i \in [m]$ has k_i possible values. Let A_i be the set of attribute i's possible values, namely $A_i = \{a_{i,1}, a_{i,2}, \ldots, a_{i,k_i}\}$. Each context is described by an ordered m-tuple $(x_1, x_2, \ldots, x_m)$, $x_i \in A_i$ for each $i \in [m]$. Ω is the Cartesian product of $A_1, A_2, \ldots, A_m$, which is the universal set of all contexts. Clearly $|\Omega| = \prod_{i=1}^{m} k_i$. The probability distribution over Ω is denoted by P. We assume it is publicly known to the auctioneer and advertisers, since it can be computed by search engine's query log and other factors. We use $P(c)$ to denote the probability of the context c and $P(B) = \sum_{c\in B} P(c)$ to denote the probability of the bundle B.

Let V be the valuation matrix. Further, $v_i(c) = V(i, c)$ is the bidder i's valuation for the context c, and $v_i(B) = \sum_{c\in B} P(c)v_i(c)$ is the bidder i's expected valuation for the bundle B. In this paper, only the optimization problem is considered. So we assume the valuation matrix is publicly known. In future, we will design an incentive mechanism for the advertisers to tell the truth.

Definition 1 A bundle B is a bracketing bundle if and only if it can be described in bracketing form: $([\cdots], [\cdots], \ldots, [\cdots])$.

For example, there are two attributes, each of which has three possible values:

- Job: engineer, teacher, student;
- Location: China, US, Canada.

Then {(teacher, China), (teacher, US)} is a bracketing bundle because it can be represented by ([teacher], [China, US]). Of course, there exist many non-bracketing bundles such as {(teacher, China), (student, US)}. It is easy to observe that the bracketing bundle is a generalized version of the attribute hiding bundle proposed in [13]. In the above example, the attribute hiding bundle (teacher, ?) equals our bracketing bundle ([teacher], [China, US, Canada]).

Definition 2 A bracketing scheme $\mathbb{S} = \{B_1, B_2, \ldots, B_t\}$ is a partition of Ω such that:

- B_i is a bracketing bundle for $1 \leqslant i \leqslant t$;
- $|B_i| \geqslant 1$ for $1 \leqslant i \leqslant t$;
- $B_i \cap B_j = \varnothing$ for i, j;
- $\bigcup_{j=1}^{t} B_j = \Omega$.

The auctioneer's expected revenue in a bracketing scheme $\mathbb{S} = \{B_1, B_2, \ldots, B_t\}$ is computed as follows:

$$revenue = \sum_{1\leqslant k\leqslant t} P(B_k) \max 2_{i\in[n]}\{\sum_{j\in B_k} \frac{P(j)}{P(B_k)} v_i(j)\}$$
$$= \sum_{1\leqslant k\leqslant t} \max 2_{i\in[n]}\{\sum_{j\in B_k} P(j)v_i(j)\}.$$

The optimal bracketing scheme is the bracketing scheme that maximizes the auctioneer's expected revenue.

In the above equation, $\max 2\{S\}$ represents the second largest value in the set S. Let $v'_i(j) = v_i(j)P(j)$. Then the revenue is $\sum_{1\leqslant k\leqslant t} \max 2_{i\in[n]}\{\sum_{j\in B_k} v'_i(j)\}$. To simplify the representation, we use $r(b)$ to denote $\max 2_{i\in[n]}\{\sum_{j\in b} v'_i(j)\}$. Thus our problem is to compute the bracketing scheme that maximizes $\sum_{1\leqslant k\leqslant t} r(B_k)$.

3 Hardness result

In this section, we will prove that computing the optimal bracketing scheme is strongly NP-hard. The proof is with the aid of Theorem 1 in [13].

Theorem 1 It is NP-hard to find the bracketing scheme that maximizes the auctioneer's revenue.

Proof First, we need to define a *multi-element bracketing scheme* $\mathbb{S}' = \{b_1, b_2, \ldots, b_s\}$. This kind of bracketing scheme only contains bundles whose sizes are greater than 1. We can find that the multi-element bracketing scheme and the original bracketing scheme are mutually convertible. For every multi-element bracketing scheme $\mathbb{S}' = \{b_1, b_2, \ldots, b_s\}$, there exists an original bracketing scheme $\mathbb{S} = \mathbb{S}' \cup \{\Omega - \cup_{1\leqslant k\leqslant s} b_k\}$. Then the revenue can be written as:

$$revenue = \sum_{1\leqslant k\leqslant s} \max 2_{i\in[n]}\{\sum_{j\in b_k} v'_i(j)\} + \sum_{c\in\Omega-\cup_{1\leqslant k\leqslant s} b_k} \max 2_{i\in[n]}\{v'_i(c)\}.$$

Let $er(b)$ represent the extra revenue of selling bundle b rather than selling contexts in b separately:

$$er(b) = \max 2_{i\in[n]}\{\sum_{j\in b} v'_i(j)\} - \sum_{c\in b} \max 2_{i\in[n]}\{v'_i(c)\}.$$

Then the auctioneer revenue is:

$$\sum_{1\leqslant k\leqslant s} er(b_k) + \sum_{c\in\Omega} \max 2_{i\in[n]}\{v'_i(c)\}.$$

It is easy to see that $\sum_{c\in\Omega} \max 2_{i\in[n]}\{v'_i(c)\}$ is constant. Thus the problem is to find a multi-element bracketing scheme $\mathbb{S}' = \{b_1, b_2, \ldots, b_s\}$ maximizes $\sum_{1\leqslant k\leqslant s} er(b_k)$.

Then the proof is similar to that of Theorem 1 in [13], which is a reduction from the Monotone one-in-three 3SAT problem. In the Monotone one-in-three 3SAT problem, there are D clauses and E variables. Further, each clause has only one positive literal. The problem determines whether there exists a satisfying assignment so that exactly one literal in each clause is set to 1. We only need to change Family (1)–(7) of the proof of Theorem 1 in [13] into below:

$$(\underline{e}, \underline{d}, [0], [0, 1], [0, 1], [0], [1], [0], [1], [0], [1], [0], [1]), \quad (1)$$

$$(\underline{e}, \underline{d}, [0, 1], [0], [0, 1], [0], [1], [0], [1], [0], [1], [0], [1]), \quad (2)$$

$$(\underline{e}, \underline{d}, [0, 1], [0, 1], [0], [0], [1], [0], [1], [0], [1], [0], [1]), \quad (3)$$

$$(\underline{e}, \underline{[0, 1]}, [0], [0], [0], [0, 1], [0, 1], [0], [1], [0], [1], [0], [1]), \quad (4)$$

$$(\underline{[0, 1]}, \underline{d}, [1], [0, 1], [0, 1], [0], [1], [0, 1], [0, 1], [0], [1], [0], [1]), \quad (5)$$

$$(\underline{[0, 1]}, \underline{d}, [0, 1], [1], [0, 1], [0], [1], [0], [1], [0, 1], [0, 1], [0], [1]), \quad (6)$$

$$(\underline{[0, 1]}, \underline{d}, [0, 1], [0, 1], [1], [0], [1], [0], [1], [0], [1], [0, 1], [0, 1]), \quad (7)$$

where $\underline{e}$ is the binary representation of integer e in bracketing form. $\underline{d}$ is the binary representation of integer d in bracketing form. $\underline{[0, 1]}$ is $[0, 1]$ repeated $\lceil \log_2(E) \rceil$ times (Family (5)–(7)) or $\lceil \log_2(D) \rceil$ times (Family (4)). □

According to [16], if a problem $\mathbb{P}$ is NP-hard and involves numbers, it is associated with two related functions: $Length[\mathbb{P}]$ and $\mathrm{Max}[\mathbb{P}]$. $Length[\mathbb{P}]$ represents a feasible encoding of the problem $\mathbb{P}$. $\mathrm{Max}[\mathbb{P}]$ represents a feasible encoding of the maximum value in the problem $\mathbb{P}$.

Proposition 1 ([16]) If a problem $\mathbb{P}$ is NP-hard and there exists a polynomial p such that $\mathrm{Max}[\mathbb{P}] \leqslant p(Length[\mathbb{P}])$, then $\mathbb{P}$ is strongly NP-hard.

Theorem 2 It is strongly NP-hard to find the bracketing scheme maximizes the auctioneer's revenue.

Proof The proof also utilizes the construction of Theorem 1 in [13]. The valuation matrix V constructed has $2(6D+E)+2$ rows and $2^{\lceil \log_2 D \rceil + \lceil \log_2 E \rceil + 11}$ columns. Let $\mathrm{Max}[\mathbb{P}]$ be the maximum value in V and $Length[\mathbb{P}]$ be the length of V. We can see

$$\mathrm{Max}[\mathbb{P}] \leqslant \max\{D, 3\} + L < D + 3 + L.$$

Since $L > \max\{D, 3\}$, let $L = D + 3$. Then we get

$$\mathrm{Max}[\mathbb{P}] < 2D + 6,$$

$$Length[\mathbb{P}] > 2^{\lceil \log_2 D \rceil + \lceil \log_2 E \rceil + 11} \cdot \log_2 L > D \cdot E \cdot \log_2 L > D,$$

when $D > 6$, $\mathrm{Max}[\mathbb{P}] \leqslant 3(Length[\mathbb{P}])$. Therefore, finding the optimal bracketing scheme is strongly NP-hard. □

4 Heuristic algorithm

Since the problem is strongly NP-hard, here we present a heuristic algorithm to deal with it.

Let $f[b]$ be the optimal revenue for selling the set b of contexts. For each singleton set b, we set $f[b] = r(b)$. The resulting revenue of this heuristic algorithm is

$$f[([a_{1,1}, a_{1,2}, \ldots, a_{1,k_1}], \ldots, [a_{m,1}, a_{m,2}, \ldots, a_{m,k_m}])].$$

The solving equation is as follows:

$$\begin{aligned} f[b] = \max\{r(b), \max_{1\leqslant j\leqslant m}\{ \max_{S_j \subseteq p(b_j)} \cdot \\ \{f[(b_1, \ldots, b_{j-1}, S_j, b_{j+1}, \ldots, b_m)] + \\ f[(b_1, \ldots, b_{j-1}, b_j \backslash S_j, b_{j+1}, \ldots, b_m)]\}\}\}. \end{aligned}$$

In the above equation, $p(b_j)$ represents the power set of the attribute set b_j except for two elements: $\varnothing$ and b_j.

The idea of our heuristic algorithm is inspired by dynamic programming. To compute $f[b]$ for a bracketing bundle $b = (b_1, b_2, \ldots, b_m)$ where b_i is the bracketing form, it is easy to see that the optimal revenue must be at least the revenue of bidding for bundle b directly. So the initial state of $f[b]$ is $r(b)$. Otherwise, we can select arbitrary attribute j and arbitrary proper subset of b_j (i.e., S_j). Then we compute the sum of $f[(b_1, \ldots, b_{j-1}, S_j, b_{j+1}, \ldots, b_m)]$ and $f[(b_1, \ldots, b_{j-1}, b_j \backslash S_j, b_{j+1}, \ldots, b_m)]$. The maximum value of the previous selection is the result of $f[b]$.

The following is the implementation of the heuristic algorithm.

Algorithm Heuristic algorithm

```
Let qb = the bracketing bundle that contains all attribute values;
Let B = the set contains all bracketing bundles;
Let B′ = the set that contains all binary-encodings of bracketing
  bundles in ascending order;
for each b in B do
    f[b] = r(b);
for each b in B′ do
    if b contains more than one context then
        for i = 1 to m do
            if attribute i contains more than one value then
                Let v = the set contains all values of attribute i;
                Let ps = the power set of v excludes empty set and v;
                for each p in ps do
                    Let dp = v − p;
                    Let b1 = the new bracketing bundle that replaces at-
  tribute i's values with p;
                    Let b2 = the new bracketing bundle that replaces at-
  tribute i's values with dp;
                    f[b] = max{f[b], f[b1] + f[b2]};
return f[qb];
```

Here we will explain the binary-encoding presentation of a bracketing bundle in detail. For example, there are three attributes:

- Gender: male, female;
- Job: teacher, engineer;
- Location: Shanghai, Beijing, Shenzhen.

So there are totally seven values. Then we use seven bits to represent these values. If a bracketing bundle contains a specific value, then the corresponding bit is set by 1. For example, if a bracketing bundle is ([male], [teacher], [Beijing, Shenzhen]), then the binary encoding is “1010011”.

The following proposition says that the algorithm may not always generate the optimal bracketing scheme. However, the experiments in Section 5 shows the algorithm has a very good performance.

Proposition 2 The heuristic algorithm may not compute the optimal bracketing scheme.

Proof Here, a counterexample will be given. There are six bidders and three attributes, each of which has two possible values:

- Attribute 1: 0, 1;
- Attribute 2: 0, 1;
- Attribute 3: 0, 1.

The valuation matrix is shown in Table 1.

Table 1 The valuation matrix

Bidder	(0,0,0)	(0,0,1)	(0,1,0)	(0,1,1)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)
1	1	0	0	0	1	0	0	0
2	0	1	0	0	0	0	0	0
3	0	0	1	1	0	0	0	0
4	0	0	0	0	0	1	0	1
5	0	0	0	0	0	0	1	0
6	1	1	1	1	1	1	1	1

It is easy to verify that $\{([0,1],0,0),(0,1,[0,1]),(1,[0,1],1),(0,0,1),(1,1,0)\}$ is the optimal bracketing scheme, while it cannot be generated by the heuristic algorithm. □

5 Experiments

In this section, based on three datasets, experiments are done to compare the optimal revenue of the bracketing scheme, the optimal revenue of the attribute hiding scheme, the revenue got by the heuristic algorithm, the revenue of the context scheme, and the optimal revenue without any scheme constraint. By comparisons, it is concluded that the optimal revenue of the bracketing scheme is close to the optimal revenue without any scheme constraint, and higher than the optimal revenue of the attribute hiding scheme and the revenue of the context scheme. What’s more, the revenue got by the heuristic algorithm is close to the optimal revenue of the bracketing scheme.

To make the experiments’ results more convinced, we use three datasets to construct the valuation matrix:

- $U(0,1)$ dataset: each value of the valuation matrix is uniformly distributed in [0, 1];
- CATS dataset: we use CATS test suite [17] to generate bidding data, and construct valuation matrix with some preprocessing;
- Yahoo! dataset: real data provided by Yahoo!.

The variables in the experiments contain the number n of bidders, the number A of attributes and the number k of values in each attribute. To simplify the experiment, we assume each attribute has the same number of values. Furthermore, given A and k, there are totally k^A contexts. Therefore the number of different partitions for k^A contexts is a Bell number $B(k^A)$ which has a faster growth rate than 2^{k^A}. In consideration of it, we fix $A = 2$ and $k = 3$, and let n increase from 3 to 15.

The experiment environment is MacBook Pro. OS is Yosemite. CPU is 2.2 GHz Intel Core i7 4 core processors. Memory is 16GB 1600MHz DDR3. Disk is 256GB SSD.

5.1 $U(0,1)$ dataset

Figure 2 illustrates the optimal revenue of the bracketing scheme, the optimal revenue of the attribute hiding scheme, the revenue got by the heuristic algorithm, the revenue of the context scheme, and the optimal revenue without any scheme constraint. For each setup, we repeat 100 times and calculate the average.

It is easy to see that:

- the revenue of the optimal bracketing scheme is better than the revenue of the optimal attribute hiding scheme;
- the revenue by the heuristic algorithm is very close to the revenue of the optimal bracketing scheme;
- the revenue of the optimal bracketing scheme is very close to the optimal revenue without any scheme constraint.

In order to observe the figure more carefully, we transform it to Table 2 where each cell represents the percentage of the optimal revenue of the corresponding scheme to the optimal revenue without any scheme constraint.

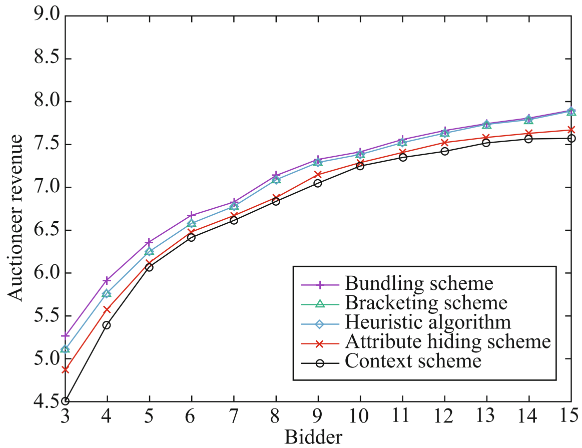


Fig. 2 The comparison of five methods' optimal revenues in $U(0,1)$ dataset

Table 2 The revenue comparison in $U(0,1)$ dataset

n	Bracketing scheme/%	Heuristic algorithm/%	Attribute hiding scheme/%	Context scheme/%
3	97	97	92	85
4	98	98	94	91
5	98	98	96	94
6	99	99	97	96
7	99	99	97	96
8	99	99	96	95
9	99	99	97	96
10	99	99	97	96
11	99	99	98	97
12	99	99	98	96
13	99	99	97	96
14	99	99	97	96
15	99	99	97	96
Average	99	99	96	94

We can see that under $U(0,1)$ dataset, the revenue of the optimal bracketing scheme is 3% higher than that of the optimal attribute hiding scheme, 5% higher than that of the context scheme, 1% lower than that of the optimal revenue without any scheme constraint.

Because the above table still can't tell the difference of performances of the optimal bracketing scheme and the heuristic algorithm, we further record the number of times that the optimal bracketing scheme equals the optimal revenue without any scheme constraint in 100 instances (Line #1 in Table 3), and the number of times that the revenue of the heuristic algorithm equals the optimal bracketing scheme in 100 instances (Line #2 in Table 3).

We can see that:

Table 3 The number of times having the same revenue in $U(0,1)$ dataset

	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
#1	10	5	15	18	26	38	49	53	56	59	67	67	71	35
#2	74	67	59	48	54	40	43	40	36	42	33	39	40	47

- with $1/3$ probability, the revenue of the optimal bracketing scheme is equal to the optimal revenue without any scheme constraint;
- with $1/2$ probability, the revenue of the heuristic algorithm is equal to the revenue of the optimal bracketing scheme.

5.2 CATS dataset

CATS (Combinatorial Auction Test Suite) [17] is a testbed for combinatorial auctions. It can simulate the real scenes of auctions. It has been widely recognized and used by academics. We use CATS 2.0 in the experiment.

Figure 3, Tables 4 and 5 are the experiment results for CATS.

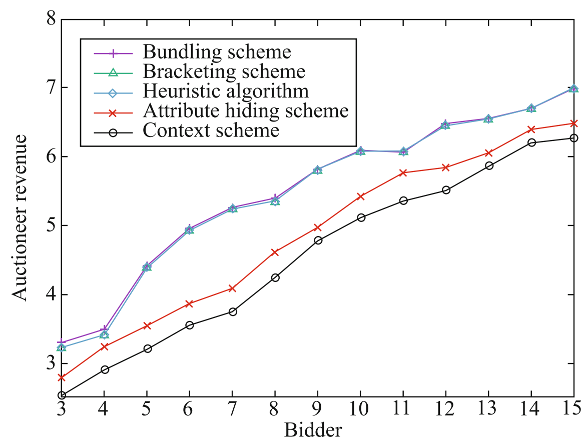


Fig. 3 The comparison of five methods' optimal revenues in CATS dataset

5.3 Yahoo! dataset

The dataset is from the open dataset of Yahoo Webscope project. It contains all bidding data of the first 1,000 keywords from June 15, 2002 to June 14, 2003. Although it has no context information, in the experiment we regard these keywords as contexts. After determine the number of attributes A and the number of values k for every attribute, we randomly select k^A contexts from the dataset.

Figure 4, Tables 6 and 7 are the experiment results for Yahoo! dataset.

5.4 Conclusions for the experiments

The experiments results on U(0,1) dataset, CATS dataset and

Table 4 The revenue comparison in CATS dataset

n	Bracketing scheme/%	Heuristic algorithm/%	Attribute hiding scheme/%	Context scheme/%
3	98	98	84	76
4	98	98	92	83
5	99	99	80	72
6	99	99	77	71
7	99	99	77	71
8	99	99	85	78
9	99	99	85	82
10	99	99	89	83
11	99	99	95	88
12	99	99	90	85
13	99	99	92	89
14	99	99	95	92
15	99	99	92	89
Average	99	99	87	81

Table 5 The number of times having the same revenue in CATS dataset

	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
#1	42	46	67	53	57	72	93	79	91	69	53	81	100	69
#2	100	79	83	58	93	79	76	95	70	74	56	67	79	78

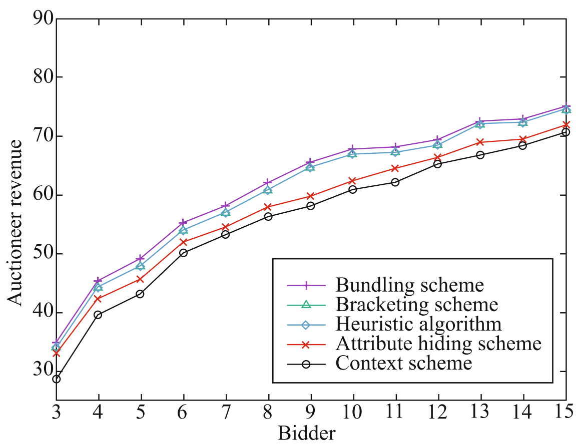


Fig. 4 The comparison of five methods' optimal revenues in Yahoo! dataset

Yahoo! dataset are consistent. Through the results, we can conclude that:

- the revenue of the optimal bracketing scheme is very close to the optimal revenue without any scheme constraint;
- the revenue of the heuristic algorithm is very close to the revenue of the optimal bracketing scheme;
- the revenue of the optimal bracketing scheme is much better than the revenue of the optimal attribute hiding scheme and the revenue of the context scheme.

6 Valuation matrix constraint

In this section, we will discuss a simplified model that every advertiser can only bid a fixed value for his favorite contexts. In other word, for each row of the valuation matrix, the non-zero cells have the same value. We will present the hardness results of two problems with this constraint.

Table 6 The revenue comparison in Yahoo! dataset

n	Bracketing scheme/%	Heuristic algorithm/%	Attribute hiding scheme/%	Context scheme/%
3	98	98	94	82
4	98	98	93	86
5	98	98	92	87
6	98	98	94	90
7	98	98	93	91
8	98	98	93	90
9	99	99	91	88
10	99	99	92	89
11	99	99	94	91
12	99	99	95	93
13	99	99	94	91
14	99	99	95	93
15	99	99	95	94
Average	99	99	93	89

Table 7 The number of times having the same revenue in Yahoo! dataset

	3	4	5	6	7	8	9	10	11	12	13	14	15	Average
#1	38	20	20	11	13	19	24	22	18	19	33	33	36	24
#2	74	69	75	65	67	59	69	66	67	69	65	71	74	68

6.1 K-anonymous signaling scheme

The K-anonymous signaling scheme means that every bundle must contain at least K contexts, which is proposed by Chen and Qin [12]. We will prove that it is NP-hard to compute the optimal K-anonymous signaling scheme with the valuation matrix constraint by a reduction from the X3C (Exact Cover By 3-Sets) problem.

Definition 3 (X3C Problem) Given a finite set $E = \{e_1, e_2, \ldots, e_{3q}\}$ and a collection $C = \{c_1, c_2, \ldots, c_m\}$ of 3-element subsets of E, determine whether there is a subcollection $C' \subseteq C$ such that every element of E is covered by exactly one member of C'.

Theorem 3 Given a valuation matrix V' such that the non-zero cells are the same for each row, integers α and K, it is NP-hard to determine whether the optimal revenue of the K-anonymous signaling scheme is at least α.

Proof We prove the problem by reducing the X3C problem to it. Let $\alpha = 3q$, $K = 4$. The valuation matrix V' has $m + 1$ rows and $4q$ columns. Each row of $2 \leqslant i \leqslant m+1$ corresponds to each element of C and each column of $q + 1 \leqslant j \leqslant 4q$

corresponds to each element of E. The valuation matrix is 0 except that:

- $V'[1, j] = 3$ for $1 \leqslant j \leqslant q$;
- for each $c_a = \{e_i, e_j, e_k\} \in C$ (suppose c_a corresponds to Row $a+1$), let $V'[a+1, i+q] = V'[a+1, j+q] = V'[a+1, k+q] = 1$.

For example, suppose $X = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ and $C = \{\{e_1, e_2, e_3\}, \{e_2, e_5, e_6\}\}$. Then the valuation matrix is:

			e_1	e_2	e_3	e_4	e_5	e_6
	3	3	0	0	0	0	0	0
$\{e_1, e_2, e_3\}$	0	0	1	1	1	0	0	0
$\{e_2, e_5, e_6\}$	0	0	0	1	0	0	1	1

We will prove there exists an exact cover if and only if the optimal revenue of 4-anonymous signaling scheme problem with the valuation matrix constraint is at least $3q$.

First, suppose that $C' = \{c'_1, c'_2, \ldots, c'_q\}$ is the exact cover. We can construct the partition $S = \{B_1, B_2, \ldots, B_q\}$ such that Bundle B_i consists of column i and the corresponding columns of each element of c'_i. So each bundle's size is 4, which satisfies the 4-anonymous signaling scheme. Clearly, the revenue of S is $3q$.

Conversely, suppose that the optimal revenue of 4-anonymous signaling scheme problem with the valuation matrix constraint is at least $3q$. Let the optimal scheme be S. Since $K = 4$, each bundle size must be at least 4. So the number of bundles in S must be at most q. We can easily see that the revenue of every bundle is at most 3 regardless of the size of the bundle. To satisfy the requirement that the revenue of S is at least $3q$, the size of S must be exactly q, which means each bundle's size must be exactly 4 and the revenue of each bundle is exactly 3. The revenue of each bundle is exactly 3 only when each bundle of S contains one column of column $1 \leqslant j \leqslant q$ whose valuation is 3 and three columns of column $q+1 \leqslant j \leqslant 4q$ (there exists one row such that the valuation of these three columns is 1). Therefore, $C' = \{S(b) | b \in S\}$, where $S(b)$ is the subset of E containing the corresponding elements of the columns ranging $q+1$ from $4q$ of b, is the exact cover. □

6.2 Cardinality constrained signaling scheme

Next, we will discuss the revenue maximization problem such that the number of bundles are fixed, which is called the cardinality constrained signaling scheme [9]. Similarly, we prove that it is NP-hard to compute the optimal revenue under the value matrix constraint by a reduction from the variant of the expected component sum problem with 0/1 entries [16].

Definition 4 (Expected Component Sum Problem with 0/1 Entries) Given a collection $C = \{v^1, v^2, \ldots, v^n\}$ of $(m-1)$-dimensional vectors $v^i = (v^i_1, v^i_2, \ldots, v^i_{m-1})$, each entry's value of which is 0 or 1, positive integers K and D, determine whether there is a K-partition $C_1, C_2, \ldots, C_K$ such that $\sum\limits_{i=1}^{K} \max\limits_{1\leqslant j\leqslant m-1} \left(\sum\limits_{v\in C_i} v_j\right) \geqslant D$.

We slightly modify the problem. We add one component of value 1 into each vector in C. Thus every vector in C is a m-dimensional vector. Then the original problem equals to determine if there is a K-partition $C_1, C_2, \ldots, C_K$ of C such that $\sum\limits_{i=1}^{K} max2_{1\leqslant j\leqslant m} \left(\sum\limits_{v\in C_i} v_j\right) \geqslant D$. Therefore, the following problem is NP-hard.

Definition 5 Given a collection $C = \{v^1, v^2, \ldots, v^n\}$ of m-dimensional vectors $v^i = (v^i_1, v^i_2, \ldots, v^i_m)$, each entry's value of which is 0 or 1 and $v^i_1 = 1$ for each $1 \leqslant i \leqslant n$, positive integers K and D, determine whether there is a K-partition $C_1, C_2, \ldots, C_K$ such that $\sum\limits_{i=1}^{K} max2_{1\leqslant j\leqslant m} \left(\sum\limits_{v\in C_i} v_j\right) \geqslant D$.

Theorem 4 Given a valuation constrained matrix V', integers α and K', it is NP-hard to determine whether the optimal revenue of the cardinality constrained signaling scheme is at least α.

Proof The proof is by a reduction from the problem defined in Definition 5. Let $\alpha = D$, $K' = K$. V' is a $m \times n$ matrix and $V'(i, j) = v^j_i$, i.e., the column i corresponds to the vector v^i.

We will prove that there exists a feasible partition $C = \{C_1, C_2, \ldots, C_K\}$ if and only if the optimal revenue of the cardinality constrained signaling scheme with valuation matrix constraint is at least D.

First, suppose that there exists a feasible partition $C = \{C_1, C_2, \ldots, C_K\}$. We construct the partition $S = \{B_1, B_2, \ldots, B_K\}$ where B_i consists of the corresponding columns of vectors in C_i. Obviously, the revenue of the partition S is at least D.

Conversely, suppose that the optimal revenue of the cardinality constrained signaling scheme is at least D. Let the optimal feasible partition be $S = \{B_1, B_2, \ldots, B_{K'}\}$ with $K' \leqslant K$. If $K' = K$, S is a feasible partition. If $K' < K$, then we do the following step until the number of bundles is K: choose a bundle whose size is greater than 1, then partition it into two bundles arbitrarily. Clearly the partition operation will not decrease the revenue. Eventually we get a partition

$S' = \{B'_1, B'_2, \ldots, B'_K\}$ whose revenue is at least D. Thus there exists a feasible partition $C = \{C_1, C_2, \ldots, C_K\}$ such that C_i contains all corresponding vectors of columns in B'_i. □

7 Conclusions

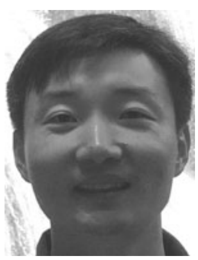
This paper analyses several bundling methods of contexts in sponsored search auctions, and proposes a new model: the bracketing method which is more flexible and natural than previous models. We prove that it is strongly NP-hard to compute the optimal bracketing scheme. We also give a heuristic algorithm and do experiments to show that our heuristic algorithm has a very good performance. Finally we prove the hardness results of computing optimal K-anonymous signaling scheme and optimal cardinality constraint signaling scheme with the valuation constraint that all non-zero cells of each row in the valuation matrix have the same value. All the results in the paper are based on the assumption that all the bidders are truthful. In future, we will design a incentive mechanism for the bidders to tell the truth.

Acknowledgements We would like to thank the anonymous reviewers for their many insightful comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 61672012).

References

1. Interactive Advertising Bureau. Iab Internet advertising revenue report 2013 full year results. Technical Report, 2014
2. Fain D C, Pedersen J O. Sponsored search: a brief history. Bulletin of the American Society for Information Science and Technology, 2006, 32(2): 12–13
3. Jansen B J, Mullen T. Sponsored search: an overview of the concept, history, and technology. International Journal of Electronic Business, 2008, 6(2): 114–131
4. Qin T, Chen W, Liu T Y. Sponsored search auctions: recent advances and future directions. ACM Transactions on Intelligent Systems and Technology, 2015, 5(4): 60
5. Even-Dar E, Kearns M, Wortman J. Sponsored search with contexts. In: Proceedings of the 3rd International Workshop on Internet and Network Economics. 2007, 312–317
6. Ghosh A, Nazerzadeh H, Sundararajan M. Computing optimal bundles for sponsored search. In: Proceedings of the 3rd International Workshop on Internet and Network Economics. 2007, 576–583
7. Emek Y, Feldman M, Gamzu I, Tennenholtz M. Revenue maximization in probabilistic single-item auctions via signaling. In: Proceedings of the 7th Ad Auctions Workshop. 2011
8. Miltersen B, Sheffet O. Send mixed signals: earn more, work less. In: Proceedings of the 13th ACM Conference on Electronic Commerce. 2012, 234–247
9. Dughmi S, Immorlica N, Roth A. Constrained signaling for welfare and revenue maximization. ACM SIGecom Exchanges, 2013, 12(1): 53–56
10. Dughmi S, Immorlica N, Roth A. Constrained signaling in auction design. In: Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms. 2014, 1341–1357
11. Dughmi S. On the hardness of signaling. In: Proceedings of the 55th IEEE Annual Symposium on Foundations of Computer Science. 2014, 354–363
12. Chen C, Qin T. k-anonymous signaling scheme. 2013, arXiv preprint arXiv:1311.6638
13. Guo M Y, Deligkas A. Revenue maximization via hiding item attributes. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. 2013
14. Papadimitriou P, Garcia-Molina H, Dasdan A, Kolay S. Output url bidding. In: Proceedings of the 37th International Conference on Very Large Data Bases. 2010, 161–172
15. Dasdan A, Santanu K, Papadimitriou P, Garcia-Molina H. Output bidding: a new search advertising model complementary to keyword bidding. In: Proceedings of the 5th Workshop on Ad Auctions. 2009
16. Garey M R, Johnson D S. Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco: Freeman. 1979
17. Leyton-Brown K, Pearson M, Shoham Y. Towards a universal test suite for combinatorial auction algorithms. In: Proceedings of the 2nd ACM Conference on Electronic Commerce. 2000, 66–76

Zheng-Dong Xia received the Bachelor's and Master's degrees in software engineering from East China Normal University, China in 2013 and 2016, respectively. His research interests include algorithm and auction theory. Now he is a developer in Tencent.



Tian-Ming Bu received the PhD degree from Fudan University, China in 1996. He is an associate professor of East China Normal University, China. His research interests include combinatorial algorithm design and algorithmic game theory.