Online Pricing for Revenue Maximization with Unknown Time Discounting Valuations

Weichao Mao[1], **Zhenzhe Zheng**[1], **Fan Wu**[1*], **Guihai Chen**[2]
[1] Shanghai Key Laboratory of Scalable Computing and Systems, Shanghai Jiao Tong University, China
[2] Department of Computer Science and Technology, Nanjing University, China
{maoweichao,zhengzhenzhe}@sjtu.edu.cn, {fwu,gchen}@cs.sjtu.edu.cn

Abstract

Online pricing mechanisms have been widely applied to resource allocation in multi-agent systems. However, most of the existing online pricing mechanisms assume buyers have fixed valuations over the time horizon, which cannot capture the dynamic nature of valuation in emerging applications. In this paper, we study the problem of revenue maximization in online auctions with unknown time discounting valuations, and model it as non-stationary multi-armed bandit optimization. We design an online pricing mechanism, namely Biased-UCB, based on unique features of the discounting valuations. We use competitive analysis to theoretically evaluate the performance guarantee of our pricing mechanism, and derive the competitive ratio. Numerical results show that our design achieves good performance in terms of revenue maximization on a real-world bidding dataset.

1 Introduction

Online pricing mechanisms have been widely adopted for allocating resources in multi-agent systems. Typical applications include cloud resource allocation [Zhang *et al.*, 2017], online advertising [Sumita *et al.*, 2017], microtask crowdsourcing [Hu and Zhang, 2017] and crowdsensed data pricing [Zheng *et al.*, 2017]. These emerging applications also require online pricing mechanisms to handle new features, one of which is time discounting valuations. For example, advertisers' willingness-to-pay usually decays over the time horizon in new mobile ad auctions, where ad space is sold in different time slots [Mehta *et al.*, 2017]. The time discounting valuation phenomenon also shows up in more general e-commerce scenarios, as customers always prefer newly released products [Chawla *et al.*, 2016].

In this paper, we study revenue maximization for posted price online auctions with unknown time discounting valuations. The items sold in the auctions can be digital goods, such as information, digital music, and software, or reusable goods, such as cloud computing processors and advertising impressions. The seller sequentially interacts with a set of buyers, and offers each buyer a price, without knowing the valuations of buyers. The buyer has time discounting valuations over the items, and decides whether to take the price by comparing her current valuation with the price. The seller's objective is to maximize the overall revenue, by setting the proper prices based on the observation of buyers' responses to prices.

Designing an online pricing mechanism for revenue maximization with time discounting valuations in an incomplete information environment has three major challenges. The first challenge is the unknown valuation setting, meaning that the seller does not know the buyers' valuations, even the valuation distributions. Most of dynamic pricing works from management science literature [Myerson, 1981; Gallego and Van Ryzin, 1994] assume that seller has the accurate knowledge of the valuation distribution. However, this assumption seldom holds in practice, as it requires a long-term marketing research and the results can still contain inaccuracies. The related works from computer science community handle this challenge by formulating the online pricing problem as a multi-armed bandit (MAB) optimization [Kleinberg and Leighton, 2003]. The intuition behind these works is to maintain a weight vector for the performance of candidate prices, and make a trade-off between exploiting the current best price and exploring the potential ultimate optimal price. However, the MAB-based pricing schemes only work for fixed valuation model. It is non-trivial to extend them to handle time discounting valuation settings.

The second challenge is the multi-dimensional private information. Since both the valuation distribution and the discounting function are unknown, a price offer may be rejected due to either the buyer originally possesses a low valuation, or the buyer's valuation goes through a large discount in the past few time slots. The traditional pricing mechanisms cannot distinguish these two cases under such limited information environments, making it difficult to determine whether to lower the price for future buyers at certain time.

The third challenge comes from the misalignment of historical and future valuations in time discounting valuation scenarios. When valuations are time discounting, historical information cannot provide accurate descriptions of future valuations, causing the revenue loss of methods that rely on historical valuations. The MAB-based pricing scheme, one representative of such methods, refers to the previous perfor-

*F. Wu is the corresponding author.

mance of prices to determine the prices for future buyers, who have significantly lower valuations than the previous buyers. These price offers inevitably get rejected by the future buyers, leading to the decrease of revenue. It is not an easy job to extend the MAB-based pricing schemes to handle the misalignment of historical information and future information, and thus new technical tools are needed.

In this paper, jointly considering the above challenges, we present an online pricing mechanism for revenue maximization in unknown time discounting valuation settings. To handle the first two challenges, we associate each candidate price with a reward distribution varying with time, and model the online pricing problem as a non-stationary MAB optimization. Specifically, we introduce an attenuation factor to attach more importance to recent intersections with buyers than transactions long ago. This procedure does not need to know the information of valuations, valuation distribution, and discounting function, but only relies on buyers' responses to prices. To address the third challenge, instead of simply feeding historical records into the weight vector, we proactively predict for future valuations. We modify the weight update scheme of candidate prices, and make our mechanism biased towards lower prices. Such biased operation amends the misalignment between historical information and future discounting valuations. Combing the above ideas, we propose the first online pricing mechanism, namely biased-UCB, for time discounting valuation setting, and derive a good competitive ratio in terms of revenue maximization.

We summarize our contributions in this paper as follows.

- First, we formulate the revenue maximization problem in online auction with time discounting valuations, based on the observations of dynamic nature of valuation in emerging applications. We model this problem as a non-stationary multi-armed bandit optimization.
- Second, we fully exploit the time discounting characteristic in valuations, and present an online pricing mechanism, namely Biased-UCB, that is biased towards lower prices. We theoretically analyze Biased-UCB, and provide the competitive ratio in the worst case, which is a function of the price discretization level and reaction time towards stepwise valuation changes.
- Finally, based on a real world real-time bidding dataset, we evaluate Biased-UCB in advertising auctions. Numerical results show that Biased-UCB outperforms the existing mechanisms in terms of revenue, and approaches to the optimal revenue.

2 Preliminaries

In this section, we present the model of posted price online auction with time discounting valuations.

A seller has unlimited supply of identical items and sequentially interacts with a set of buyers. The time horizon is divided into T slots, and is denoted by a set $\mathbb{T} = \{1, 2, \ldots, T\}$. In each slot t, one buyer b_i shows up and requests one copy of the item. We use $g(t)$ to denote the inherent discounting function for the "quality" of the item over the time horizon. For example, in mobile ad auctions, the click-through rate of the ad space decreases with time [Mehta *et al.*, 2017]; in data marketplace, the accuracy of the data decays over time [Zheng *et al.*, 2017]. Since buyers may have different responses to the quality decrease of the item, we represent buyer b_i's *discounted valuation* at slot t as

$$v_i(t) = \max\{v_i \cdot d_i(g(t)), 1\} \tag{1}$$

where $d_i(g(t))$ is the *discounting function* for buyer b_i, and v_i is the *original valuation*, denoting the valuation for the item with the highest quality. We consider the independent identical valuation case, in which the original valuations of buyers follow the same distribution with cumulative distribution function $F(x)$. For the convenience of analysis, we normalize buyers' original valuations into the range of $[1, \bar{v}]$. Since $d_i(g(t))$ is also a function of t, we simply use $d_i(t)$ to denote the discounting function of buyer b_i in the following discussion. A discounting function can be any decreasing function subject to $d_i(t) \in (0, 1]$ for all $t \in \mathbb{T}$. We assume that different discounting functions are bounded by a constant, i.e., $\frac{d_i(t)}{d_j(t)} \leq \eta$ for all i, j and t. This assumption is based on the observation that the buyers usually have different but similar perspectives over the quality of the item, making the discounting functions not far away from each other. We denote $d(t) = \min_i d_i(t)$. Following traditional prior-independent mechanism design [Goldberg *et al.*, 2001], we assume that both the cumulative distribution function and discounting functions are unknown to the seller.

At each slot t, the seller offers a price p_t to the current buyer b_i. We restrict the candidate prices to be discrete. Specifically, at each slot t, we allow the seller to select a price from a discrete candidate price vector $\hat{\boldsymbol{p}} = (\hat{p}_0, \hat{p}_1, \cdots, \hat{p}_H)$, where $\hat{p}_k = (1+\beta)^k$ for any $0 \leq k \leq H$ and $\beta > 0$. Since v_i is normalized into $[1, \bar{v}]$ and $d_i(t)$'s are upper bounded by 1, we have $H = \lfloor \log_{1+\beta} \bar{v} \rfloor$. The pricing strategies with discrete prices will bear a loss of the revenue by at most a $(1+\beta)$ factor, and thus β can be regarded as a trade-off between optimality and computational complexity. Bidder b_i will decide whether to take the offer by comparing the price p_t with her current discounted valuation $v_i(t)$. Based on the decision, the *utility* function of buyer b_i can be expressed as

$$u_i = \begin{cases} v_i(t) - p_t, & \text{if } b_i \text{ accepts the offer,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

As buyers are rational, buyer b_i takes the offer if and only if $v_i(t) \geq p_t$. The seller's objective is to maximize the *revenue*, which is defined as the sum of all accepted prices along the time horizon.

We emphasize two important features of the above posted price online auction model. First, the seller would not learn the valuation $v_i(t)$ of any buyer, but only observes buyers' responses to the prices. Second, in each time slot, the seller has to decide the price for the current buyer before seeing the next buyer. In an online auction, the seller has to utilize a pricing mechanism that can gradually learn the optimal prices through interacting with the buyers.

Following [Goldberg *et al.*, 2001], we adopt the concept of competitive analysis to measure the performance of different pricing mechanisms. If the cumulative distribution function

Algorithm 1: DescendUponRejection

```
k ← H;
for t ← 1 to T do
    Offer price p̂_k to the buyer at slot t;
    if price p̂_k is rejected then
        k ← k − 1;
```

F and discounting function $d_i(t)$'s are known, the best strategy for the seller is to compute

$$p_t^* = \arg\max_{p\in[1,\bar{v}]} p \cdot F(1 - \frac{p}{d_i(t)}) \tag{3}$$

and then offer price p_t^* to the buyer at slot t. Following [Kleinberg and Leighton, 2003], we call this strategy the *ex ante* optimal strategy, and call the revenue obtained by this strategy the *ex ante* optimal revenue, in that it only observes the distribution F, but not the individual realization of actual v_i's. We will utilize this revenue as a benchmark to evaluate the performance of different pricing mechanisms.

3 Online Pricing Mechanism Design

In this section, we present the pricing mechanism to maximize revenue in online auction with time discounting valuations. We begin with a simple setting, where all buyers' original valuations are fixed as a constant value. We provide a trivial pricing strategy that can perfectly fit in this setting. We further consider the general setting, where the original valuations are random variables following the same distribution, and propose a new pricing mechanism, namely Biased-UCB.

3.1 A Simple Case: Fixed Original Valuations

In this case, $v_i(t)$ is non-increasing in t since all v_i's are equal to an unknown constant v^* and $d_i(t)$'s are non-increasing.

For convenience of analysis, we introduce the notation of *segment*. We divide the whole time horizon $\mathbb{T}$ into $H+1$ segments $\mathbb{S} = \{S_0, S_1, \cdots, S_H\}$, where S_i is the set of slots in which $\hat{p}_i$ is the optimal discrete price regarding to the lower bound of buyers' discounted valuations. Formally,

$$S_i = \{t \in \mathbb{T} \mid \hat{p}_i \le v^* \cdot d(t) < \hat{p}_{i+1}\}. \tag{4}$$

Please note that one or more segments may be empty if $v^* < \hat{p}_H$. In the following, we only refer to the non-empty segments when using the term segment. We assume that every segment is sufficiently long, such that $|S_i| \ge 2(H+1)$ for all $S_i \in \mathbb{S}$. Note that we can always make this inequality hold true by choosing an appropriate value for β.

We now provide a naive pricing strategy, *DescendUponRejection* in Algorithm 1. This strategy performs well in the fixed valuation setting, and is $\frac{1}{2\eta}$-competitive to the discrete *ex ante* optimal strategy even from worst-case analysis. *DescendUponRejection* makes only one "error guess" in each segment, except for the very first segment, where it may spend up to $H+1$ slots seeking the optimal price.

The effectiveness of *DescendUponRejection* relies on the fact that the discounted valuation $v_i(t)$'s are non-increasing in t. Nevertheless, its performance can be arbitrarily bad in the general setting, where the original valuations are not constant. The reason is it cannot tell whether the price is rejected by a small v_i from F, or by a sharp discount in $d(t)$. For example, consider the non-discounting case where $d(t) = 1$ for all $t \in \mathbb{T}$, and the v_i's are drawn from a Gaussian distribution. Every time this naive strategy is rejected by a small v_i, it will descend to a lower price, even though the valuation is actually not discounting in time. Soon enough, this strategy will end up offering only the lowest price.

Algorithm 2: Biased-UCB

```
Initialize u_{i,t} ← 0, n_{i,t} ← 0 for all t ∈ T, 0 ≤ i ≤ H;
for t ← 1 to H + 1 do
    Offer price p̂_{t−1} at slot t;
    UpdateWeight(t − 1, t);
for t ← H + 2 to T do
    for i ← 0 to H do
        n̂_{i,t−1} ← Σ_{s=1}^{t−1} n_{i,s};
        m̂_{i,t−1} ← Σ_{s=1}^{t−1} γ^{t−s} n_{i,s};
        û_{i,t−1} ← (Σ_{s=1}^{t−1} γ^{t−s} u_{i,s}) / m̂_{i,t−1};
        w_{i,t−1} ← û_{i,t−1} + sqrt(c·ln(t−1) / n̂_{i,t−1});
    k ← arg max_{0≤i≤H} w_{i,t−1};
    Offer price p̂_k at slot t;
    UpdateWeight(k, t);
```

Algorithm 3: UpdateWeight

```
Input: Two integers k and t, indicating price p̂_k is
       offered at slot t
if Price p̂_k is accepted at slot t then
    for i ← 0 to k do
        u_{i,t} ← p̂_i;
        n_{i,t} ← 1;
else
    for i ← k to H do
        u_{i,t} ← 0;
        n_{i,t} ← 1;
```

3.2 General Case: Random Original Valuations

In the general case where v_i's are random variables, we propose the Biased-UCB algorithm in Algorithm 2. Parameter $u_{i,t}$ denotes the profit made at slot t by offering price $\hat{p}_i$ to the buyer. Parameter $n_{i,t}$ reflects whether price $\hat{p}_i$ has been *tried* at slot t. We have $u_{i,t} = \hat{p}_i$, $n_{i,t} = 1$ if price $\hat{p}_i$ is tried and accepted at slot t, $u_{i,t} = 0, n_{i,t} = 1$ if price $\hat{p}_i$ is tried but rejected, and $u_{i,t} = 0, n_{i,t} = 0$ if price $\hat{p}_i$ is not tried. Parameter c determines the trade-off between the exploitation and exploration. γ is the *attenuation factor* that measures to what extent the historical information is valued, where $0 < \gamma \le 1$. The $UpdateWeight$ procedure is defined in Algorithm 3.

The Biased-UCB algorithm follows the Upper Confidence Bound (UCB) framework from bandit problems. In the classical UCB1 algorithm proposed in [Auer *et al.*, 2002], the

authors keep a record of the average reward ($\hat{u}_{i,t}$) of each arm, and use the number of plays ($\hat{n}_{i,t}$) to denote the uncertainty of the arms. In our problem, however, since the reward distribution behind each candidate price is not fixed, we are confronted with a non-stationary MAB problem, and thus we value the recent information more than the historical records a long time ago. In this case, we utilize the parameter γ to make the value of information attenuate over time.

Additionally, in $UpdateWeight$, instead of only updating the weight of one particular price $\hat{p}_k$, we update the weights of all the prices no higher than (or no lower than) this offered price. Here, we distinguish the expression of *trying* a candidate price from *offering* a candidate price. Only one price will be offered to the buyer at one slot, but we can imagine the results of trying some other prices. For example, if $\hat{p}_k$ is offered and accepted at a certain slot, we are sure that all the prices no higher than $\hat{p}_k$ will also be accepted. We then hypothetically try these prices and also update their information accordingly (line 3 and 4). On the other hand, if $\hat{p}_k$ is rejected, the profit information of all the prices no lower than $\hat{p}_k$ will be updated with 0 (line 7 and 8).

Our proposed algorithm is *biased*, in that it always encourages lower prices and suppresses higher prices. In the following subsection, we will see this biased characteristic fits perfectly in the discounting setting, as well as provides much convenience for mathematical analysis.

3.3 Theoretical Analysis of Biased-UCB

In [Kleinberg and Leighton, 2003], the authors provide mathematical analysis for applying UCB1 algorithm to posted-price online auctions. However, the techniques they employed cannot be applied to our non-stationary MAB setting. One of the fundamental challenges in our problem is that the reward distribution behind each non-stationary arm is not constant, and thus the Chernoff-Hoeffding bound is not available. Therefore, we have to carry out our theoretical analysis from a brand-new perspective.

We will first analyze the performance of Biased-UCB in the fixed original valuation case as defined in 3.1, and then extend the proof idea to the general case. In the following discussion, we focus on the discounting function lower bound $d(t)$, and by doing so we bear a loss of revenue by no more than a factor η.

We propose three properties, namely, *accurate start*, *stable optimality*, and *quick reaction*. Accurate start requires an algorithm to find exactly the optimal price at the very beginning; stable optimality guarantees the algorithm will stick to the optimal price once it is found; and quick reaction asks the algorithm to find the new optimal price quickly once the optimal price changes. These three properties together ensure the pricing mechanism is competitive.

We show Biased-UCB indeed possesses these properties, and its competitive ratio towards discrete *ex ante* optimal strategy is lower bounded by $\frac{1}{\eta}(1-\beta r)(1-\frac{1}{1+\beta r})$, where $r=\min(\lfloor\frac{\sqrt{2}-1}{\beta}\rfloor, H+1)$.

Accurate Start

Recall that in 3.1, one severe drawback of *DescendUponRejection* is that it may waste a long time in seeking the optimal price in the first segment. To address this issue, we want a competitive pricing strategy to "start" in a reasonably fast way in the first segment. Since the UCB framework forces our algorithm to try each price at least once in the first $H+1$ slots, we put the accurate start constraint on the $(H+2)$-th slot by proving the following theorem.

Theorem 3.1. *At the $(H+2)$-th slot, Biased-UCB always offers the optimal price of the first segment.*

Proof. Let S_e denote the first segment, where e not necessarily equals H. According to our previous definitions, $\hat{p}_e$ denotes the optimal price of segment S_e, i.e., $\hat{p}_e \le v^* \cdot d(t) < \hat{p}_{e+1}$ for all $t \in S_e$. It's then equivalent to prove that $w_{e,H+1} > w_{k,H+1}$ for all $k \in \{0,1,\dots,e-1,e+1,\dots,H\}$.

For any candidate price $\hat{p}_{k_1}(e < k_1 \le H)$, we must have $\hat{u}_{k_1,H+1}=0$ and $\hat{n}_{k_1,H+1}=k-e$, since every time a price lower than $\hat{p}_{k_1}$ is offered and rejected, $\hat{p}_{k_1}$ will also be hypothetically tried and rejected. Similarly, for any candidate price $\hat{p}_{k_2}(0 \le k_2 \le e)$, we must have $\hat{n}_{k_2,H+1}=e-k_2+1$ and $\hat{u}_{k_2,H+1}=\frac{\sum_{s=1}^{H+1}\gamma^{t-s}u_{k_2,s}}{\sum_{s=1}^{H+1}\gamma^{t-s}n_{k_2,s}}=(1+\beta)^{k_2}$. Therefore, the following two inequalities trivially hold true:

- For $e < k_1 \le H$, $w_{e,H+1}=(1+\beta)^e+\sqrt{\frac{c\cdot\ln(H+1)}{1}} > 0+\sqrt{\frac{c\cdot\ln(H+1)}{k_1-e}}=w_{k_1,H+1}$.
- For $0 \le k_2 < e$, $w_{e,H+1}=(1+\beta)^e+\sqrt{\frac{c\cdot\ln(H+1)}{1}} > (1+\beta)^{k_2}+\sqrt{\frac{c\cdot\ln(H+1)}{e-k_2+1}}=w_{k_2,H+1}$.

□

Stable Optimality

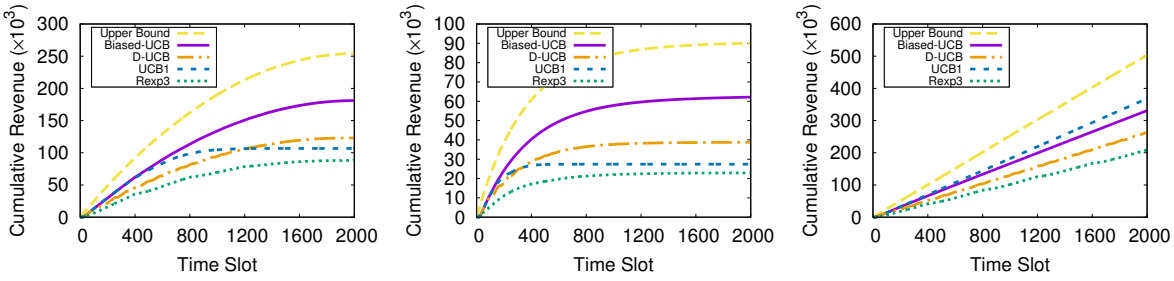
The stable optimality property requires that once an algorithm finds the optimal price of a segment, it will stick to that price until the end of this segment. We show Biased-UCB partially possesses this property by proving the following theorem.

Theorem 3.2. *For the first r segments, once Biased-UCB offers the optimal price of this segment at a certain slot, it will continue offering that optimal price in the following slots of this segment. By abandoning the last $e-r+1$ segments and the first Δ slots in each remaining segment, the competitive ratio towards discrete ex ante optimal revenue is lower bounded by $\frac{1}{\eta}(1-\beta r)(1-\frac{1}{1+\beta r})$, where $r=\min(\lfloor\frac{\sqrt{2}-1}{\beta}\rfloor, H+1)$.*

Proof. We leave the proof of Theorem 3.2 to our technical report [Mao *et al.*, 2018] due to limitation of space. □

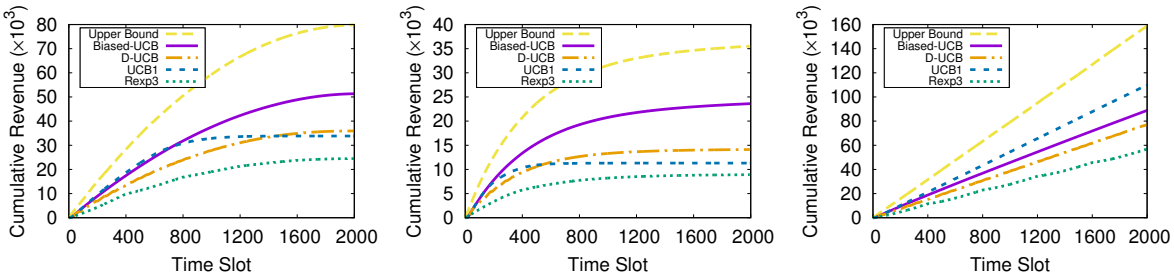
Quick Reaction

According to the stable optimality property, $\hat{p}_h$ is the offered price by the end of segment S_h. When first entering segment S_{h-1}, where price $\hat{p}_h$ will be rejected and produce 0 profit, it may take some time for our algorithm to realize $\hat{p}_h$ is not optimal anymore and adjust the weights accordingly. The quick reaction property requires that our algorithm should spend no more than a bounded number (Δ) of slots seeking the new optimal price $\hat{p}_{h-1}$. We show Biased-UCB partially possesses this property by proving the following theorem.



(a) Linear (b) Exponential (c) Non-discounting

Figure 1: Cumulative revenue of different mechanisms.



(a) Linear (b) Exponential (c) Non-discounting

Figure 2: Cumulative revenue for dispersed valuations.

Theorem 3.3. *For the first r segments, it takes Biased-UCB no more than Δ slots before switching to the new optimal price when entering a new segment. The value of r and the competitive ratio remain the same as in Theorem 3.2.*

Proof. We leave the proof of Theorem 3.3 to our technical report [Mao *et al.*, 2018] due to limitation of space. □

Now we extend the preceding proof idea to the general case. Since buyers' original valuations are not fixed in the general case, our previous definition of segment no longer holds. We present a slightly different definition of *segment* based on the price p_t^* offered by the discrete *ex ante* optimal strategy at slot t.

$$S_i = \{t \in \mathbb{T} \mid p_t^* = \hat{p}_i\}. \tag{5}$$

In the general case, the validity of the three properties relies on a further assumption that buyers' original valuations are not too dispersed. Formally, we assume $\bar{v} \leq \frac{d(t)}{d(t+\delta)}$ for all $t \in \mathbb{T}$, where the value of δ determines the balance between the stringency of the cumulative distribution function F and optimality of the guaranteed revenue. Intuitively, this assumption ensures that violations of the three properties can only occur near the two endpoints of a segment.

Following the same procedure as the simple case, we abandon the first $\delta + \Delta$ slots and the last δ slots in each segment, and argue that the remaining slots satisfy the three properties. Therefore, Biased-UCB is at least $\frac{1}{\eta}[1 - \beta r - \frac{2\delta(e+1)}{T}](1 - \frac{1}{1+\beta r})$-competitive towards discrete *ex ante* optimal strategy in the general case, where $r = \min(\lfloor \frac{\sqrt{2}-1}{\beta} \rfloor, H+1)$, assuming $\bar{v} \leq \frac{d(t)}{d(t+\delta)}$ for all $t \in \mathbb{T}$. Please note that the performance guarantee relies on an overly stringent assumption on buyers' valuations and seems relatively weak. This is because we are performing worst-case analysis without putting any restriction on the discounting functions. Considering the difficulty of our general discounting model, we think this relatively weak bound is acceptable.

4 Numerical Results

In this section, we empirically compare our mechanism with the upper bound of total revenue, and with mechanisms adapted from existing bandit algorithms, including UCB1 [Auer *et al.*, 2002], D-UCB [Garivier and Moulines, 2011] and Rexp3 [Besbes *et al.*, 2014]. The upper bound of total revenue is defined as the sum of all buyers' discounted valuations, and is obviously the upper bound of any pricing mechanism. D-UCB is a non-stationary MAB algorithm that employs similar concept to the attenuation factor in our paper. Rexp3 is also a non-stationary MAB algorithm. It divides the time horizon into several batches and restarts an Exp3 algorithm [Auer *et al.*, 1995] at the beginning of each batch.

We use the real-world bidding feedback log [Zhang *et al.*, 2014] from the iPinYou company as our dataset. This dataset was released in a real-time bidding competition held by iPinYou in 2013. It contains logs of ad auctions, bids, impressions, clicks and final conversions during ten days in 2013, and we use the 9.58×10^6 records of bidding prices on June 6, 2013 as our valuation distribution. The bidding prices

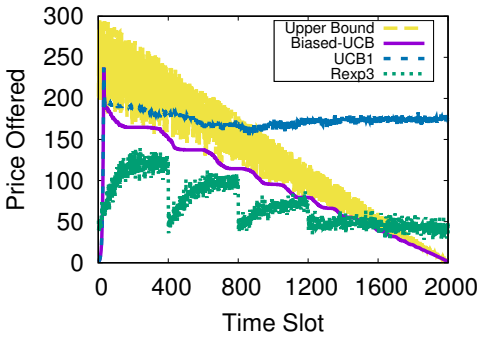


Figure 3: Prices offered at each time slot.

range from 227 to 300 and the unit is RMB fen $\times 1000$.

We try three discounting functions, including linear discount $g(t) = 1 - \frac{t-1}{T}$, exponential discount $g(t) = \alpha^{1-t}$ and non-discount $g(t) = 1$. We set $d_i(g(t)) = a_i \cdot g(t)$, where a_i's are independently and uniformly drawn from $[0.8, 1.0]$. In each run we randomly select 2000 bids (i.e., $T = 2000$) from the dataset as buyers' original valuations, and multiply each price by the current value of discounting function to get the discounted valuation. We set price discretization level $\beta = 0.2$, exploration-exploitation control parameter $\tau = 0.5$ for Rexp3 and $c = 400.0$ for other algorithms, exponential discount factor $\alpha = 1.003$, attenuation factor $\gamma = 0.9$ for Biased-UCB and D-UCB, and batch size $\Delta = 400$ for Rexp3. All results are averaged over 200 runs.

Figure 1 shows the evaluation results for cumulative revenue obtained in the first t slots. We can see that Biased-UCB performs better than existing methods. The total revenue of Biased-UCB is 71.2% of the upper bound for linear discount, and 69.0% for exponential discount. In the non-discount case, Biased-UCB performs slightly worse than UCB1, since it is designed to try lower prices first when being rejected. Nevertheless, it still achieves 89.9% revenue of UCB1.

Figure 2 shows the results for more dispersed valuations, where the original valuations are drawn from a Gaussian distribution $\mathcal{N}(150, 30^2)$ rather than the bidding dataset. Although our theoretical analysis relies on the compactness of buyers' valuation distribution, we can see Biased-UCB still performs well on very dispersed valuations.

To give an intuitive description of different mechanisms, we now plot the prices offered by different mechanisms at each time slot. Take linear discount as an example. The prices offered by the upper bound are exactly buyers' actual discounted valuations. These prices roughly form a triangular shape in Figure 3, and at the first slot, buyers' discounted valuations are in the range of $[0.8 \times 227, 1.0 \times 300]$. The UCB1 algorithm is a stationary MAB algorithm. It first finds an optimal price at the early stage, accumulating (overly) large weight on that price, and then stick to that price ever since. Therefore, the prices offered by UCB1 basically form a horizontal line in Figure 3. The prices offered by Rexp3, unsurprisingly, show an obvious restarting pattern. The D-UCB algorithm possesses a similar restarting pattern as Rexp3, only with shorter restarting period, and is thus omitted. The prices offered by the Biased-UCB mechanism firmly follow the lower bound of buyers' discounted valuations, and thus Biased-UCB achieves good performance in terms of revenue.

5 Related Works

In [Lavi and Nisan, 2000], the problem of online auction was first introduced to the literature of computer science. Later, Goldberg *et al.* [2001] began the study of (offline) auctions for digital goods. Bar-Yossef *et al.* [2002] studied online auctions for digital goods, and employed randomization to ensure truthfulness. Online learning was first applied to online auctions in an early version of [Blum *et al.*, 2004]. Kleinberg and Leighton [2003] demonstrated how to apply bandit algorithms to online auctions. Hajiaghayi *el al.* [2005] studied the problem of online scheduling for reusable goods. However, these mechanisms did not take time discounting valuation into consideration.

In management science literature, the stochastic demand model is categorized into dynamic pricing problems. Gallego and Van Ryzin [1994] investigated the problem of selling a given stock of items by a deadline. Problems of similar setting were also considered in [Levin *et al.*, 2010; Gershkov *et al.*, 2017]. Nevertheless, these works are only concerned with finite inventories, assuming the demand curve is known to the seller, and are essentially different from our problem. Mechanisms with discounting values are also considered in computer science literature. Secretary problems with weights and discounts were discussed in [Babaioff *et al.*, 2009]. Wu *et al.* [2014] presented a strategy-proof online auction with discounting valuations, but they assume the discounting functions are known to the seller, and their objective is to maximize social welfare instead of revenue. A recent work [Xu *et al.*, 2017] considered dynamic pricing with time-variant rewards, but the variant part in their setting is the utility function instead of the valuations, and their weight function is basically linear.

In the seminal work of [Auer *et al.*, 2002], the UCB framework was proposed to solve multi-armed bandit problems. For bandits with non-stationary rewards, Besbes *et al.* [2014] suggested dividing the time horizon into batches, and restarting a traditional bandit algorithm at the beginning of each batch. Bandits with abruptly changing rewards were discussed in [Hartland *et al.*, 2006; Garivier and Moulines, 2011]. Similar works include bandit problem with Markovian rewards [Tekin and Liu, 2010] and reward functions following Brownian motion [Slivkins and Upfal, 2008]. Nonetheless, bandit algorithms need to be carefully modified before being applied to pricing problems.

6 Conclusion

In this paper, we have studied the problem of revenue maximization in posted-price auctions with unknown time discounting valuations. We have modeled the revenue maximization problem as a non-stationary MAB optimization, and proposed the Biased-UCB mechanism based on unique features of the discounting valuations. We have theoretically analyzed the lower bound of the competitive ratio. Our numerical results have shown that our design achieves good performance in terms of revenue.

Acknowledgments

This work was supported in part by the State Key Development Program for Basic Research of China (973 project 2014CB340303), in part by China NSF grant 61672348, 61672353, and 61472252, and in part by Shanghai Science and Technology fund 15220721300. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies or the government.

References

[Auer *et al.*, 1995] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *FOCS*, 1995.

[Auer *et al.*, 2002] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[Babaioff *et al.*, 2009] Moshe Babaioff, Michael Dinitz, Anupam Gupta, Nicole Immorlica, and Kunal Talwar. Secretary problems: weights and discounts. In *SODA*, 2009.

[Bar-Yossef *et al.*, 2002] Ziv Bar-Yossef, Kirsten Hildrum, and Felix Wu. Incentive-compatible online auctions for digital goods. In *SODA*, 2002.

[Besbes *et al.*, 2014] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed-bandit problem with non-stationary rewards. In *NIPS*, 2014.

[Blum *et al.*, 2004] Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. *Theoretical Computer Science*, 324(2-3):137–146, 2004.

[Chawla *et al.*, 2016] Shuchi Chawla, Nikhil R Devanur, Anna R Karlin, and Balasubranianian Sivan. Simple pricing schemes for consumers with evolving values. In *SODA*, 2016.

[Gallego and Van Ryzin, 1994] Guillermo Gallego and Garrett Van Ryzin. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Science*, 40(8):999–1020, 1994.

[Garivier and Moulines, 2011] Aurélien Garivier and Eric Moulines. On upper-confidence bound policies for switching bandit problems. In *ALT*, 2011.

[Gershkov *et al.*, 2017] Alex Gershkov, Benny Moldovanu, and Philipp Strack. Revenue-maximizing mechanisms with strategic customers and unknown, markovian demand. *Management Science*, 2017.

[Goldberg *et al.*, 2001] Andrew V Goldberg, Jason D Hartline, and Andrew Wright. Competitive auctions and digital goods. In *SODA*, 2001.

[Hajiaghayi, 2005] Mohammad T Hajiaghayi. Online auctions with re-usable goods. In *EC*, 2005.

[Hartland *et al.*, 2006] Cédric Hartland, Sylvain Gelly, Nicolas Baskiotis, Olivier Teytaud, and Michele Sebag. Multi-armed bandit, dynamic environments and meta-bandits. *Online Trading between Exploration and Exploitation Workshop, NIPS*, 2006.

[Hu and Zhang, 2017] Zehong Hu and Jie Zhang. Optimal posted-price mechanism in microtask crowdsourcing. In *IJCAI*, 2017.

[Kleinberg and Leighton, 2003] Robert Kleinberg and Tom Leighton. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *FOCS*, 2003.

[Lavi and Nisan, 2000] Ron Lavi and Noam Nisan. Competitive analysis of incentive compatible on-line auctions. In *EC*, 2000.

[Levin *et al.*, 2010] Yuri Levin, Jeff McGill, and Mikhail Nediak. Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. *Production and Operations Management*, 19(1):40–60, 2010.

[Mao *et al.*, 2018] Weichao Mao, Zhenzhe Zheng, Fan Wu, and Guihai Chen. Technical report, 2018. https://drive.google.com/open?id=18GwsahidHGFGxhrwMqCAhgexpJKNRA2l.

[Mehta *et al.*, 2017] Sameer Mehta, Milind Dawande, Ganesh Janakiraman, and Vijay Mookerjee. Sustaining a good impression: mechanisms for selling 'partitioned' impressions at ad-exchanges. 2017.

[Myerson, 1981] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[Slivkins and Upfal, 2008] Aleksandrs Slivkins and Eli Upfal. Adapting to a changing environment: the brownian restless bandits. In *COLT*, 2008.

[Sumita *et al.*, 2017] Hanna Sumita, Yasushi Kawase, Sumio Fujita, and Takuro Fukunaga. Online optimization of video-ad allocation. In *IJCAI*, 2017.

[Tekin and Liu, 2010] Cem Tekin and Mingyan Liu. Online algorithms for the multi-armed bandit problem with markovian rewards. In *Allerton*, 2010.

[Wu *et al.*, 2014] Fan Wu, Junming Liu, Zhenzhe Zheng, and Guihai Chen. A strategy-proof online auction with time discounting values. In *AAAI*, 2014.

[Xu *et al.*, 2017] Lei Xu, Chunxiao Jiang, Yi Qian, Youjian Zhao, Jianhua Li, and Yong Ren. Dynamic privacy pricing: A multi-armed bandit approach with time-variant rewards. *IEEE Transactions on Information Forensics and Security*, 12(2):271–285, 2017.

[Zhang *et al.*, 2014] Weinan Zhang, Shuai Yuan, Jun Wang, and Xuehua Shen. Real-time bidding benchmarking with ipinyou dataset. *arXiv preprint arXiv:1407.7073*, 2014.

[Zhang *et al.*, 2017] Zijun Zhang, Zongpeng Li, and Chuan Wu. Optimal posted prices for online cloud resource allocation. In *SIGMETRICS*, 2017.

[Zheng *et al.*, 2017] Zhenzhe Zheng, Yanqing Peng, Fan Wu, Shaojie Tang, and Guihai Chen. An online pricing mechanism for mobile crowdsensing data markets. In *MobiHoc*, 2017.