Value of Learning in Sponsored Search Auctions

Sai-Ming Li[1], Mohammad Mahdian[1], and R. Preston McAfee[1]

Yahoo! Inc.
Santa Clara, CA, USA.

Abstract. The standard business model in the sponsored search marketplace is to sell click-throughs to the advertisers. This involves running an auction that allocates advertisement opportunities based on the value the advertiser is willing to pay per click, times the click-through rate of the advertiser. The click-through rate of an advertiser is the probability that if their ad is shown, it would be clicked on by the user. This quantity is unknown in advance, and is learned using historical click data about the advertiser. In this paper, we first show that in an auction that does not explore enough to discover the click-through rate of the ads, an advertiser has an incentive to increase their bid by an amount that we call *value of learning.* This means that in sponsored search auctions, exploration is necessary not only to improve the efficiency (a subject which has been studied in the machine learning literature), but also to improve the incentive properties of the mechanism. Secondly, we show through an intuitive theoretical argument as well as extensive simulations that a mechanism that sorts ads based on their expected value per impression *plus* their value of learning, increases the revenue *even in the short term.*

1 Introduction

Online advertising provides the major revenue source for most online services today. The most common standard in the online advertising marketplace is Pay-Per-Click, which means that the publisher sells "click-throughs" to the advertisers. An advertiser is charged only when a user clicks on their ad. The allocation and pricing of such ads are often done through an auction: each advertiser specifies the maximum they are willing to pay for a click-through, and the auction mechanism decides which ad(s) should be shown and how much each of them should pay in the event of a click. The most prominent example of online ad auctions is sponsored search auctions, which allocate the ad space on the side of search results pages of major search engines.

The efficient allocation of ad space in a pay-per-click system is based on the expected value from each impression of the ad. This expected value is the product of the advertiser's value for each click and the probability that if the ad is shown, it will be clicked on. Estimating the latter parameter, called the *Click-Through Rate* (CTR), is a central piece of an ad allocation engine.

The problem of efficiently allocating the ad space and simultaneously estimating the CTR for future is essentially a form of the multi-armed bandits problem [4]. In this problem, the task is to strike a balance between *exploring*, i.e., showing an ad to get a better estimate of its CTR, and *exploiting*, i.e., showing ads that have the best performance, according to our current estimates of the CTRs. There are several papers that give explore-exploit algorithms for this problem from a machine learning perspective [1, 6, 7, 14–16, 11]. The goal of this paper is not to give yet another explore-exploit algorithm for sponsored search (even though our analysis involves designing an algorithm for a simple setting). Instead, we seek to make two points in this paper: First, even a second-price auction, which is incentive compatible in most settings, fails to be incentive compatible when the mechanism does not perform exploration. Specifically, in such a mechanism, an advertiser has an incentive to increase their bid by some amount, which we call their *value of learning*. This means that performing exploration improves not only the efficiency of the mechanism, but also its incentive properties. Furthermore, this suggests an exploration-exploitation mechanism that is quite natural from an economic standpoint: sort the ads based on their expected value per impression *plus* their value of learning. Multi-armed bandits algorithms based on Upper Confidence Bounds [3, 4] can be considered in this vein.

Second, despite the intuition that "exploration has some short-term cost", we show that incorporating value of learning in the auction mechanism (the way described above) can lead to a higher revenue *even in the short term.* In other words, in a mechanism that performs exploration by incorporating value of learning, the cost of learning is paid by the advertisers, and not by the seller. This is based on the intuition that value of learning gives higher boost to advertisers in lower slots, thereby helping to level the playing field among advertisers competing for the same ad space and increasing the competition. We show this through a non-rigorous theoretical argument (as making the statement rigorous requires arguing about a complex Bayesian model), as well as extensive simulations using real advertisers' data.

Previous Work. In addition to the vast literature on the explore-exploit algroithms for various forms of the multi-armed bandit problem [1, 6, 7, 14–16, 11], the paper by Goel and Munagala [10] is related to our work. They attack the problem of uncertainty about click-through rates using a different approach, by allowing the advertiser to make a per-impression as well as a per-click bid.

2 Model and Notations

We consider a setting where n advertisers (or bidders) are competing to be placed in one of the m slots, numbered 1 through m. Advertiser i has a value v_i and bids b_i for a click. We assume a separable model for click-through rates, i.e., there is a value γ_j associated with each slot $j = 1, \ldots, m$ (called the position bias of slot j) and a value λ_i for each advertiser (called the clickability of this advertiser), such that if the ad of advertiser i is displayed in position j, it will be clicked on with probability $\gamma_j \lambda_i$.[1] We assume that the slots are numbered

[1] This is assumed to be independent of other ads placed on the page; for models that do not make this assumption see [9, 2, 12].

in decreasing order of their position bias, i.e., $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_m$. The exact clickability of advertisers are not known, and the system tries to estimate these quantities based on the past performance of the ad. We denote the estimate of the clickability of advertiser i by $\hat{\lambda}_i$. Note that this value can change as time progresses.

The most common mechanism for allocating the ad slots to the advertisers is the so called *generalized second price auction (GSP)* [8, 17]. In this mechanism, the advertisers are ordered in their decreasing order of their $\hat{\lambda}_i b_i$, and slots 1 through m are allocated to the top m advertisers in this order (or are left empty if the number of advertisers is less than m). The amount the i'th advertiser is charged in the event of a click is the minimum this advertiser could bid and still win the same slot. This means that if advertiser i is allocated slot i, the price per click for this advertiser is $p_i := \hat{\lambda}_{i+1} b_{i+1} / \hat{\lambda}_i$. As shown in [8, 17], bidding truthfully is not an equilibrium in the GSP mechanism, i.e., the advertisers have incentive to submit bids other than their true value per click, but it has full-information equilibria which coincide with the outcome of the Vickrey-Clark-Groves (VCG) mechanism, which is a well-knwon incentive compatible mechanism. In the case that there is only one slot ($m = 1$), the GSP mechanism is the same as the second-price auction, which is incentive compatible. In the next section, we will focus on this case to separate out the strategic issues of the GSP mechanism from the incentive issues resulting from the uncertainty in click-through rates.

3 Incentives in auctions without exploration

In this section, we look at the second price auction described in the previous section from the perspective of one advertiser, and show that in a repeated second-price auction without exploration, when there is uncertainty about the clickability of the advertiser, it is no longer in the advertiser's best interest to bid her value per click. Specifically, the advertiser has the incentive to increase her bid in order to induce the mechanism to explore her. This is done through a simple model defined below.

We assume that the advertiser i faces a price per impression distribution $\mathcal{D}_p$, i.e., the highest bid times click-through rate among other advertisers is distributed according to $\mathcal{D}_p$. We make the simplifying assumption that this distribution does not change over time and is independent in each time step. The advertiser has a value per click v_i, and a clickability λ_i, which is distributed according to a prior $\mathcal{D}_\lambda$. Neither the advertiser nor the auctioneer knows the value of λ_i. Instead, an unbiased estimate $\hat{\lambda}_i$ of this value is calculated using Bayesian updating given the current history, i.e., the estimate $\hat{\lambda}_i$ at any point in time is equal to the expected value of λ_i given the prior $\mathcal{D}_\lambda$ and the observed click/no-click history. In each time step, the advertiser decides how much to bid (the bid b_i can change as time progresses); then a price p is picked according to $\mathcal{D}_p$, and if $\hat{\lambda}_i \geq p$, i's ad is displayed and i is charged $p/\hat{\lambda}_i$ in the event of a click. Both advertiser i and the auctioneer will observe whether or not the ad is clicked on. We assume an infinite time horizon (i.e., an infinite sequence of auctions) and a discount factor of $\delta < 1$.

In the above model, if there is no uncertainty about the clickability λ_i (i.e., if $\mathcal{D}_\lambda$ has a singleton support), the optimal strategy for advertiser i to bid $b_i = v_i$ in every round. In the rest of this section, we show that this is not the case in general where there is uncertainty about λ_i. To demonstrate this point, we calculate the advertiser's optimal strategy as a recurrence and prove that it is non-negative in general. We will also give a lower bound for the advertiser's optimal bid in the case of uniform distributions.

At any point, the state can be described by two numbers (k, N), indicating a state where the ad of the advertiser has been shown N times and out of these impressions, k of them have lead to clicks. Based on the prior $\mathcal{D}_\lambda$, the posterior distribution of the clickability at this state can be computed. Let $\hat{\lambda}_{k,N}$ denote the expected value of this posterior distribution. Let $U(k, N)$ denote the optimal utility of an infinite sequence of auctions, starting from this posterior distribution on λ. We obtain a recurrence relation for U as follows: let b denote the bid of the advertiser in the first round. If $p < \hat{\lambda}_{k,N}b$, then the advertiser wins and has to pay $p/\hat{\lambda}_{k,N}$ in the event of a click. By the definition of $\hat{\lambda}$, this means that the advertiser pays p per impression in expectation. Therefore, the total utility of the advertiser in this round can be written as:

$$\Pr[p < \hat{\lambda}_{k,N}b](v\hat{\lambda}_{k,N} - E[p|p < \hat{\lambda}_{k,N}b]).$$

We denote the above value by $g(\hat{\lambda}_{k,N}, b)$. Denoting the pdf and the cdf of $\mathcal{D}_p$ by $f(.)$ and $F(.)$ respectively, the above expression can be written as:

$$g(\hat{\lambda}_{k,N}, b) = \int_0^{\hat{\lambda}_{k,N}b} (v\hat{\lambda}_{k,N} - p)f(p)dp. \tag{1}$$

If the ad is shown (which happens with probability $F(\hat{\lambda}_{k,N}b)$), it is either clicked on (with probability $\hat{\lambda}_{k,N}$), or not (with probability $1 - \hat{\lambda}_{k,N}$); leading us to one of the states $(k+1, N+1)$ or $(k, N+1)$. Therefore, the overall utility of the advertiser can be written as:

$$\begin{aligned} U(k, N) = \max_b \Bigg\{ & g(\hat{\lambda}_{k,N}, b) \\ & + \delta F(\hat{\lambda}_{k,N}b)\Big(\hat{\lambda}_{k,N}U(k+1, N+1) + (1 - \hat{\lambda}_{k,N})U(k, N+1)\Big) \\ & + \delta(1 - F(\hat{\lambda}_{k,N}b))U(k, N)\Bigg\}. \end{aligned} \tag{2}$$

This implies:

$$U(k, N) = \frac{1}{1-\delta} \max_b \Bigg\{ g(\hat{\lambda}_{k,N}, b) + \delta F(\hat{\lambda}_{k,N}b)\Delta(k, N)\Bigg\}, \tag{3}$$

where

$$\varDelta(k,N) := \hat{\lambda}_{k,N}U(k+1,N+1) + (1-\hat{\lambda}_{k,N})U(k,N+1) - U(k,N). \qquad (4)$$

Intuitively, $\varDelta(k,N)$ indicates the advertiser's value for the information she obtains by observing the outcome of one additional impression. We take the derivative of the expression in Equation (3) with respect to $z = \hat{\lambda}_{k,n}b$ to compute the optimal bid. Using (1), this derivative can be written as:

$$\begin{aligned}\frac{\partial(g(\hat{\lambda}_{k,N},b) + \delta F(\hat{\lambda}_{k,N}b)\varDelta(k,N))}{\partial z} &= (v\hat{\lambda}_{k,N} - z)f(z) + \delta f(z)\varDelta(k,N),\\ &= f(z)(v\hat{\lambda}_{k,N} + \delta\varDelta(k,N) - z).\end{aligned}$$

Given that $f(z)$ is non-negative, the root of the linear term in the paranthesis satisfies the second-order condition and is therefore a maximizer of the function. Thus, the optimal bid of the advertiser can be written as:

$$b^* = v + \frac{\delta}{\hat{\lambda}_{k,N}}\varDelta(k,N). \qquad (5)$$

This shows that the optimal bid of the advertiser is not the true value per click v, but the value per click plus some additional term. This additional term is proportional to the information value of one additional impression, and can be expressed with a recurrence relation. In general, this recurrence is hard to solve explicitly. However, here we prove that the optimal bid of the advertiser is always greater than or equal to her value per click. Later, we will give a lower bound on the optimal bid in the special case of uniform distributions.

Theorem 1. *In the above model of repeated auctions, the optimal bid of the advertiser in every state (k,N) is at least v.*

Proof. By Equation (5), we need to prove that $\varDelta(k,N) \geq 0$. In other words, we need to show that the expected optimal revenue starting from the state (k,N) (which we call scenario 1) is less than the optimal revenue when we first start from (k,N), observe the outcome of one impression, and then proceed (we call this scenario 2).[2] We prove this inequality by analyzing the strategy for scenario 2 that simulates the optimal strategy of scenario 1. This gives a lower bound on the optimal strategy in scenario 2.

To simulate the optimal strategy of scenario 1 in scenario 2, in each step we take the optimal bid b of scenario 1, and submit a bid in scenario 2 that leads

[2] Note that this statement is not trivial, since the additional information (the outcome of one impression) is observed by both the advertiser and the auctioneer. While the additional information enables the advertiser to make more informed decisions to improve her utility, it also enables the auctioneer to allocate and price future impressions more accurately. It is not clear a priori whether the latter effect helps or hurts the advertiser.

to the same expected bid per impression. For example, in the first step (i.e., when we are in state (k, N) in scenario 1), the corresponding bid in scenario 2 is either $b\hat{\lambda}_{k,N}/\hat{\lambda}_{k+1,N+1}$ or $b\hat{\lambda}_{k,N}/\hat{\lambda}_{k,N+1}$, depending on whether the state is $(k+1, N+1)$ or $(k, N+1)$. The expected utility of the advertiser in scenario 2 in this step can be written as

$$\hat{\lambda}_{k,N} g(\hat{\lambda}_{k+1,N+1}, b\hat{\lambda}_{k,N}/\hat{\lambda}_{k+1,N+1}) + (1 - \hat{\lambda}_{k,N}) g(\hat{\lambda}_{k,N+1}, b\hat{\lambda}_{k,N}/\hat{\lambda}_{k+1,N+1})$$

Using (1), this can be written as:

$$\hat{\lambda}_{k,N} \int_0^{\hat{\lambda}_{k,N} b} (v\hat{\lambda}_{k+1,N+1} - p) f(p) dp + (1 - \hat{\lambda}_{k,N}) \int_0^{\hat{\lambda}_{k,N} b} (v\hat{\lambda}_{k,N+1} - p) f(p) dp$$
$$= \int_0^{\hat{\lambda}_{k,N} b} \big(v(\hat{\lambda}_{k,N}\hat{\lambda}_{k+1,N+1} + (1 - \hat{\lambda}_{k,N})\hat{\lambda}_{k,N+1}) - p \big) f(p) dp$$

Using the definition $\hat{\lambda}_{k,N}$ as the posterior probability of getting a click conditioned on having had k clicks out of the first N impressions, it is easy to show that

$$\hat{\lambda}_{k,N}\hat{\lambda}_{k+1,N+1} + (1 - \hat{\lambda}_{k,N})\hat{\lambda}_{k,N+1} = \hat{\lambda}_{k,N}. \tag{6}$$

Therefore, the expected utility in the first step in scenario 2 is equal to

$$\int_0^{\hat{\lambda}_{k,N} b} (v\hat{\lambda}_{k,N} - p) f(p) dp = g(\hat{\lambda}_{k,N}, b),$$

which is the same as the expected utility in the first step in scenario 1. Similarly, in any step the simulated strategy in scenario 2 obtains the same expected payoff as in scenario 1. Thus, $\Delta(k, N) \geq 0$.

The above theorem only shows that the optimal bid of the advertiser is never smaller than her true value. To show that this bid is sometimes strictly larger than the value, we focus on the case of uniform distributions: We assume a uniform prior $\mathcal{D}_\lambda = U[0, 1]$ on the clickability and a uniform price distribution $\mathcal{D}_p = U[0, 1]$. Straightforward calculations using the Bayes rule and the prior $\mathcal{D}_\lambda$ shows that the posterior probability density for the clickability λ in a state (k, N) is

$$(n+1)\binom{n}{k}\lambda^k (1-\lambda)^{N-k}.$$

The expected value of λ given this posterior is $\hat{\lambda}_{k,N} = \frac{k+1}{N+2}$, and the function $g(.)$ from (1) can be written as $g(\hat{\lambda}_{k,N}, b) = \hat{\lambda}_{k,N}^2 b(v - \frac{b}{2})$.

Theorem 2. *In the above model of repeated auction, the optimal bid of the advertiser in every state (k, N) is strictly larger than v. More specifically, we have $\Delta(k, N) = \Omega(N^{-2})$.*

Proof (Proof Sketch). As in the proof of Theorem 1, we need to bound the difference between the optimal expected utility of scenarios 1 and 2. Again, we do this by taking the optimal strategy in scenario 1, and simulating it in scenario 2. Unlike the proof of Theorem 1, we simulate a strategy that submits a bid of b in scenario 1 by submitting the same bid in scenario 2. First, notice that with this strategy, the probability of winning the first auction in scenario 2 can be written as

$$\hat{\lambda}_{k,N}F(b\hat{\lambda}_{k+1,N+1})+(1-\hat{\lambda}_{k,N})F(b\hat{\lambda}_{k,N+1}) = (\hat{\lambda}_{k,N}\hat{\lambda}_{k+1,N+1}+(1-\hat{\lambda}_{k,N})\hat{\lambda}_{k,N+1})b$$

Using (6), the above probability is equal to $\hat{\lambda}_{k,N}b$, which is the same as the probability of winning in scenario 1. This ensures that the simulated strategy in scenario 2 has the same branching probabilities as the optimal strategy in scenario 1. Next, we need to bound the difference between the expected utility of one auction in the two scenarios. Here we only do this for the first auction. The inequality for the other auctions can be proved similarly. The difference between the expected utilities of the advertiser in the first auction in the two scenarios can be written as:

$$\begin{aligned}
&\hat{\lambda}_{k,N}g(\hat{\lambda}_{k+1,N+1},b) + (1-\hat{\lambda}_{k,N})g(\hat{\lambda}_{k,N+1},b) - g(\hat{\lambda}_{k,N},b)\\
&= b(v-\frac{b}{2})\left(\hat{\lambda}_{k,N}\hat{\lambda}^2_{k+1,N+1} + (1-\hat{\lambda}_{k,N})\hat{\lambda}^2_{k,N+1} - \hat{\lambda}^2_{k,N}\right)\\
&= b(v-\frac{b}{2})\left((\frac{k+1}{N+2})(\frac{k+2}{N+3})^2 + (1-\frac{k+1}{N+2})(\frac{k+1}{N+3})^2 - (\frac{k+1}{N+2})^2\right)\\
&= b(v-\frac{b}{2})\frac{(k+1)(N-k+1)}{(N+2)^2(N+3)^2} = \Omega(N^{-2}).
\end{aligned}$$

4 Value of learning

Given the result in the previous section, we can define the *value of learning* of an advertiser as the difference between the optimal bid of the advertiser and her value-per-click. More formally, the value of learning is the difference between the Gittins index in the Markov Decision Process (MDP) defined based on the auction. If we could compute these indices, we could simply design an alternative auction mechanism that allocates according to these indices, thereby achieving the optimal MDP solution and eliminating the incentive to overbid. Unfortunately, Gittins indices are quite hard to compute.

As a practical alternative, we can use proxies for the value of learning that are easy to compute. Perhaps the simplest method for doing this is to take the value of learning of an advertiser to be proportional to the variance of our estimate of the clickability of this advertiser. This has the advantage that it can be easily computed, and gives a boost to ads that we currently do not have an accurate estimate of its clickability.

The strongest theoretical evidence that taking the value of learning of an ad to be proportional to the variance of its clickability estimate and then sorting the ads based on their expected value per impression plus their value of learning leads to close-to-optimal outcomes comes from the literature on the multi-armed bandits problem. Multi-armed bandits algorithms based Upper Confidence Bounds are shown to achieve asymptotically optimal regrets [3, 4]. These algorithms in each iteration pick the arm that has the maximum expected value plus an additional factor that is close to the variance of the performance of the arm so far. The literature on multi-armed bandits is a vast literature and we do not intend to add yet another algorithm to this literature. Instead, we describe a practical method for incorporating the value of learning in sponsored search auctions, and analyze its revenue and efficiency impacts through simulations with real advertisers' bid and click-through rate data.

A practical value-of-learning mechanism. Recall that in sponsored search, a sequence of m slots need to be allocated to the advertisers. The position bias of slot j is denoted by γ_j. At any point in time, the history for each ad consists of the number of times this advertiser is shown in each slot, and the number of such instances that have lead to clicks. We can compute the *cumulative expected clicks* ec_i of advertiser i as the sum of the position biases of the positions this ad is shown so far. This is essentially the number of clicks we would expect this ad to receive, if it had a clickability of 1. Our estimate of the clickability is then

$$\hat{\lambda}_i = \frac{c_i}{ec_i}, \tag{7}$$

where c_i is the total number of clicks advertiser i has received. It is not hard to show that in a reasonable Bayesian setting (e.g., uniform priors), the variance of this estimate is of the order of $\sqrt{\frac{\hat{\lambda}_i}{ec_i}}$. Therefore, we define the value of learning for this advertiser as $\hat{\theta}_i b_i$, where

$$\hat{\theta}_i = C\sqrt{\frac{\hat{\lambda}_i}{ec_i}} \tag{8}$$

for a constant C. We will change the value of C in our simulations to study the effects of increaseing the value of learning on the efficiency of and revenue of the auctions. The mechanism computes a score s_i for each advertiser as follows:

$$s_i = b_i(\hat{\lambda}_i + \hat{\theta}_i). \tag{9}$$

It then sorts the advertisers in decreasing order of their scores, allocates the i'th position to the i'th advertiser in this order, and in the event of a click, charges this advertiser an amount equal to

$$p_i = \frac{b_{i+1}(\hat{\lambda}_{i+1} + \hat{\theta}_{i+1})}{(\hat{\lambda}_i + \hat{\theta}_i)} \tag{10}$$

Note that this value is never greater than the bid of the advertiser.

5 Revenue of auctions with value of learning

Intuitively, one might think that exploration in repeated sponsored search auctions is a costly activity that is done in order to achieve a better outcome in the long run. In fact, many of the exploration-exploitation algorithms based on the ϵ-greedy algorithm for the multi-armed bandits problem give out exploration impressions to the advertisers for free [7]. However, we will show experimentally in the next section that the mechanism in the previous section can lead to a higher revenue *even in the short term.* In this section, we explain the theoretical intuition behind this result.

In auction theory [13], it is known that giving an advantage to weaker bidders (e.g., minority-owned firms participating in spectrum auctions [5]) can increase the revenue by leveling the playing field between competing bidders. Here, also, the value of learning added to each advertiser's bid is inversely proportional to the square root of the number of times this ad has been clicked on. This means that an ad that is typically in a lower position has a higher value of learning, and this can increase the price that the advertisers in higher positions pay. A formal proof of this fact in the model with repeated auctions is out of reach, as it would require analyzing optimal strategies in a Bayesian multi-player version of the model studied in Section 3. Instead, we ignore incentives resulting from learning by studying a one-shot auction, and then prove that the GSP-like mechanism that allocates slots to bidders in decreasing order of $(\lambda_i + \theta_i)v_i$ has a minimal envy-free equilibrium similar to the VCG-equivalent equilibria of [8, 17]. Furthermore, the revenue of this equilibrium at the λ_i, θ_i values computed in the previous section is typically larger than the similar revenue when θ_i's are zero. The result, whose proof is omitted here, can be stated as follows.

Theorem 3. *Consider a multi-slot auction between n bidders. Assume that the i'th bidder has a value of $v_i\hat{\lambda}_i\gamma_j$ for being placed in slot j. The mechanism $\mathcal{M}_\theta$ sorts the advertisers based on their $(\hat{\lambda}_i + \theta_i)b_i$, allocates the i'th slot to the i'th advertiser in this order (which we call advertiser i), and charges her $\frac{(\hat{\lambda}_{i+1}+\theta_{i+1})b_{i+1}\hat{\lambda}_i\gamma_i}{\hat{\lambda}_i+\theta_i}$ in expectation. This mechanism has a minimal envy-free equilibrium whose revenue is denoted by $R(\theta)$. Furthermore, let $\hat{\lambda}_i$ and $\hat{\theta}_i$ be the values calculated in* (7) *and* (8) *and assume that the ordering of the values $(\hat{\lambda}_i+\hat{\theta}_i)v_i$ is the same as the ordering of the values $\hat{\lambda}_i v_i$ and that the historical number of clicks c_i of a bidder in a higher slot is higher. Then we have $R(\hat{\theta}) > R(0)$.*

The main assumption of this theorem (apart from restricting equilibrium analysis to a 1-shot game) is that the ordering of the advertisers in decreasing order of $(\hat{\lambda}_i + \hat{\theta}_i)v_i$ is the same as their ordering in decreasing order of $\hat{\lambda}_i v_i$, and their ordering in decreasing order of c_i. Since θ_i's are typically small and higher slots get more clicks, this assumption is often true, except for rare cases where the mechanism reverses the ordering to do some exploration. The above theorem guarantees that in *normal* cases, the mechanism with value of learning has a higher revenue. Intuitively, this revenue increase can more than make up for the occasional revenue loss due to exploration. This is why in the simulations in the next section we will see that incorporating value of learning leads to a considerable increase in revenue, averaged over thousands of auctions.

6 Simulation Results

In this section we provide simulation results to illustrate the performance of incorporating value of learning in the auction mechanism when applied in a popular search engine like Yahoo! Search. We collect a representative sample of sponsored search results from the Yahoo! search log. For each search sample, we collect the position bias for each position due to the specific page layout used. For each ad, we also collect the bid and its estimated clickability at the time of sampling.

For the purpose of conducting the simulation study, we assume that the ads in each search in the dataset are unique, *i.e.*, the same ad cannot appear across multiple sample searches, hence its clickability estimate only depend on its own history. We also assume that the page layout remains the same, *i.e.*, the same position bias as in the search log will be used to simulate click event and efficiency. We use the estimated clickability of each ad at the time of sampling as their true clickability. The simulation is initialized by simulating a small number of impressions using the assumed true clickability of each ad, and the position effect is based on the one at position one. Then the initial clickability estimate of each ad is computed based on the simulated clicks during those impressions.

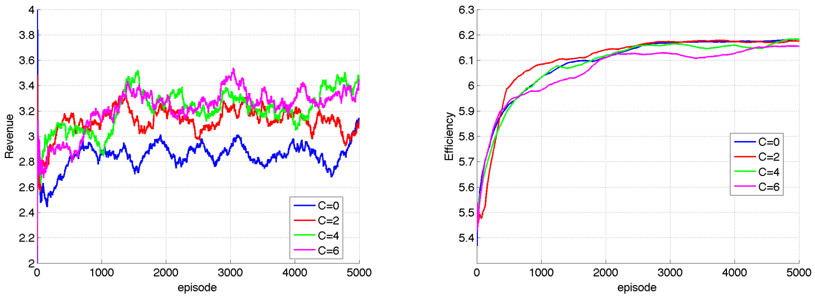
After the initialization stage, we simulated the sample searches for 5,000 episodes. Each episode involves simulating all sample searches once. For each sample search s, the value of learning term $\hat{\theta}_{s,i}$ for ad i was determined based on the current clickability estimate $\hat{\lambda}_{s,i}$ and cumulative expected clicks as in (8). The price and rank of each ad was determined by the GSP algorithm using the ranking score (9) and pricing equation (10). The number of clicks for each ad was simulated using the probability of click $\lambda_{s,i}\gamma_{s,j}$, where j is the slot occupied by ad i. Then we updated the clickability estimate for all ads after every simulated search according to (7). After each episode, we computed the total revenue and the total efficiency across the sample searches based on the simulated clicks, the PPC of the ads that were clicked, and their bids. Specifically, the revenue R and efficiency E at each episode is defined as

$$R = \sum_{s,i} p_{s,i} c_{s,i}, \qquad E = \sum_{s,i} b_{s,i} \lambda_{s,i} \gamma_{s,j},$$

where $p_{s,i}$, $c_{s,i}$, $b_{s,i}$, $\lambda_{s,i}$, denote the price per click, number of clicks, bid, and clickability of ad i in search s respectively, and $\gamma_{s,j}$ denote the position effect of the slot (j) occupied by ad i in search s. Note that in computing the efficiency, we made the simplifying assumption that the bid $b_{s,i}$ does not change over time, and it is the same as the value per click for the advertiser. Nevertheless, we believe that E serves as a good proxy for the true efficiency of the algorithm under investigation.

We simulated the auction and click behavior for a range of C to illustrate the effect of imposing different degree of learning in the mechanism. The case $C = 0$ corresponds to the case when there is no value of learning included in the auction. The higher C is, the more impact value of learning has on price and ranking, and hence revenue and efficiency. Figure 1 (a) shows the moving average of the total revenue generated over the duration of the simulation, and figure 1 (b) shows the moving average of the total efficiency. The moving average window used in the graphs has width 400. As can be seen in the figures, the revenue is consistently higher when value of learning is used in the auction. Furthermore, efficiency is higher in the transient when the appropriate value of learning ($C = 2$) is used. It should be noticed that when C is large ($C = 6$), both the transient and final efficiency can suffer as too much exploration is being done.

We also simulated the case when not all ads in each sample search are shown in every auction by reducing the number of slots that can be shown in each auction m to five (in Yahoo! search this can be as high as twelve). In other words, when the number of ads available is more than five, the algorithm is forced to select only five ads to show, with the rest not getting any exposure at all. As can be seen in Figure 2, the power of incorporating value of learning in the auction is more evident in this case. The efficiency of the auction with C other than zero is much higher than when C is zero. This is because the set of ads that are shown are fixed very early as other ads with high clickability are never given a chance to prove themselves. The revenue is also higher when C is non-zero, due to the price effect of value of learning as well as improved efficiency.



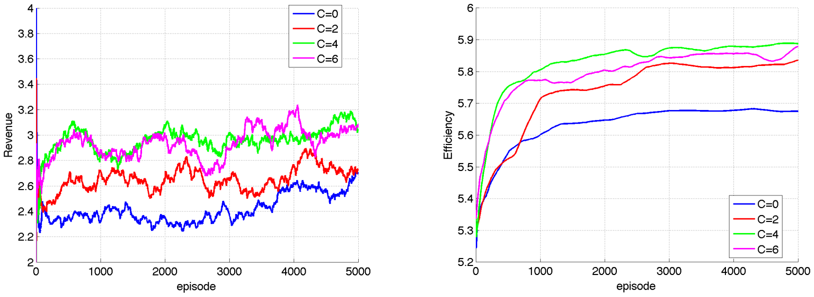
(a) revenue

(b) efficiency

Fig. 1. Moving average of revenue and efficiency for different setting of C.

References

1. D. Agarwal, B. Chen, and P. Elango. Explore/exploit schemes for web content optimization. In *ICDM*, 2009.



(a) revenue

(b) efficiency

Fig. 2. Moving average of revenue and efficiency for different setting of C when $m = 5$.

2. Gagan Aggarwal, Jon Feldman, S. Muthukrishnan, and Martin Pal. Sponsored Search Auctions with Markovian Users. *Workshop on Ad Auctions*, 2008.
3. Jean Yves Audibert, Rémi Munos, and Csaba Szepesvári. Exploration–Exploitation Tradeoff Using Variance Estimates in Multi-armed Bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.
4. P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2-3):235–256, 2002.
5. Ian Ayres and Peter Cramton. Deficit reduction through diversity: How affirmative action at the fcc increased auction competition. *Stanford Law Review*, 48(4):761–815, 1996.
6. Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing Truthful Multi-Armed Bandit Mechanisms. *EC*, pages 79–88, 2009.
7. Nikhil R. Devanur and Sham M. Kakade. The Price of Truthfulness for Pay-Per-Click Auctions. *EC*, pages 99–106, 2009.
8. Ben Edelman, Michael Ostrovsky, and Michael Schwarz. Internet Advertising and the Generalized Second-price Auction: Selling Billions of Dollars Worth of Keywords. *American Economic Review*, 97(1):242–259, 2007.
9. A. Ghosh and M. Mahdian. Externalities in Online Advertising. *WWW*, pages 161–168, 2008.
10. Ashish Goel and Kamesh Munagala. Hybrid keyword search auctions. In *Proceedings of the 18th World Wide Web conference*, 2009.
11. Satyen Kale, Mohammad Mahdian, Kunal Punera, Tamas Sarlos, and Aneesh Sharma. Position-aware multi-armed bandits. working paper, 2010.
12. David Kempe and Mohammad Mahdian. A Cascade Model for Externalities in Sponsored Search. *Workshop on Internet Ad Auctions*, 2008.
13. Vijay Krishna. *Auction Theory*. Academic Press, 2002.
14. J. Langford, L. Li, Y. Vorobeychik, and J. Wortman. Maintaining Equilibria during Exploration in Sponsored Search Auctions. *LNCS*, 4858:119, 2007.
15. Sandeep Pandey and Christopher Olston. Handling Advertisements of Unknown Quality in Search Advertising. *NIPS*, 19:1065, 2007.
16. A. Das Sarma, S. Gujar, and Y. Narahari. Multi-Armed Bandit Mechanisms for Multi-Slot Sponsored Search Auctions. *Arxiv preprint arXiv:1001.1414*, 2010.
17. Hal R. Varian. Position Auctions. *International Journal of Industrial Organization*, 25(6):1163–1178, 2007.