Sponsored Search Auctions with Markovian Users

Gagan Aggarwal, Jon Feldman, S. Muthukrishnan, and Martin Pál

Google, Inc.
76 Ninth Avenue, 4th Floor, New York, NY, 10011
1600 Amphitheatre Pkwy, Mountain View, CA, 94043
{gagana,jonfeld,muthu,mpal}@google.com

Abstract. Sponsored search involves running an auction among advertisers who bid in order to have their ad shown next to search results for specific keywords. The most popular auction for sponsored search is the "Generalized Second Price" (GSP) auction where advertisers are assigned to slots in the decreasing order of their *score*, which is defined as the product of their bid and click-through rate. One of the main advantages of this simple ranking is that bidding strategy is intuitive: to move up to a more prominent slot on the results page, bid more. This makes it simple for advertisers to strategize. However this ranking only maximizes efficiency under the assumption that the probability of a user clicking on an ad is independent of the other ads shown on the page. We study a Markovian user model that does not make this assumption. Under this model, the most efficient assignment is no longer a simple ranking function as in GSP. We show that the optimal assignment can be found efficiently (even in near-linear time). As a result of the more sophisticated structure of the optimal assignment, bidding dynamics become more complex: indeed it is no longer clear that bidding more moves one higher on the page. Our main technical result is that despite the added complexity of the bidding dynamics, the optimal assignment has the property that ad position is still monotone in bid. Thus even in this richer user model, our mechanism retains the core bidding dynamics of the GSP auction that make it useful for advertisers.

1 Introduction

Targeted advertisements on search queries is an increasingly important advertising medium, attracting large numbers of advertisers and users. When a user poses a query, the search engine returns search results together with advertisements that are placed into positions, usually arranged linearly down the page, top to bottom. On most major search engines, the assignment of ads to positions is determined by an auction among all advertisers who placed a bid on a keyword that matches the query. The user might click on one or more of the ads, in which case (in the pay-per-click model) the advertiser receiving the click pays the search engine a price determined by the auction.

C. Papadimitriou and S. Zhang (Eds.): WINE 2008, LNCS 5385, pp. 621–628, 2008.

In the past few years, the sponsored search model has been highly successful commercially, and the research community is attempting to understand the underlying dynamics, explain the behavior of the market and to improve the auction algorithms. The most common auction being run today is the *Generalized Second Price* (GSP) auction: Each bidder i submits a bid b_i stating the maximum amount they are willing to pay for a click, and the bidders are placed in descending order of $b_i p_i$, where p_i is what is called the *click-through-rate* of advertiser i; i.e., the probability that a user will click on the ad, given that the user looks at it. Much of previous research on sponsored search auctions has fixed this sort order, and focused on understanding the implications of different pricing schemes, assuming strategic behavior on the part of the advertisers. We now know something about GSP's equilibrium properties [8,17,3], alternative pricing that will make it truthful [3], and to some extent, impact on the revenue in principle [8] and via simulations [16].

However, by fixing this sort order, and assuming that an ad's clicks are independent of the other ads on the page, prior work exogenizes an important third party in sponsored search: the *search engine user*. Unfortunately, there is very little guidance on this in the literature, even though the user's behavior is the essential ingredient that defines *the commodity* the advertisers are bidding on, and its value.

We suggest a different framework for principled understanding of sponsored search auctions:

– Define a suitable probabilistic model for search engine user behavior upon being presented the ads.
– Once this model is fixed, ask the traditional mechanism design questions of how to assign the ads to slots, and how to price them.
– Analyze the given mechanism from the perspective of the bidders (e.g., strategies) and the search engine (e.g., user satisfaction, efficiency, revenue).

There are certain well-accepted observations about the user's interaction with the sponsored search ads that should inform the model:

– The higher the ad is on the page, the more users see it and thus click on it.
– The "better" the ad is, the more users click on it, where the ad's "goodness" depends on its inherent quality, and how well it matches the user query.

These properties govern not only how the auction is run but also how advertisers think about their bidding strategy: they prefer to appear higher and get more clicks. Indeed even though GSP is not truthful under a private value model (where each bidder has some inherent private value v_i for a click), its ranking function makes bidding strategy simple: to obtain a higher (more prominent) slot, bid higher and/or make your ad better. This simplicity is very important, since an advertiser may not have a precise notion of click value on which to base their bid. Indeed some advertisers are hoping to generate traffic just to attract attention to their brand, and the value of this attention is less clear than for an advertiser who is making direct sales through the Internet.

In this paper, we propose a natural Markov model for user clicks that retains the properties above, and no longer assumes that the number of clicks an ad receives is independent of the other ads on the page. We show that simple ranking no longer finds the optimal assignment of ads to slots under this model, but the optimal assignment can still be found efficiently (in near-linear time). Our main technical result is to show that under this new assignment algorithm, the auction mechanism still has a simple bidding strategy; i.e., we prove that a bidder's ad slot is monotone in her bid.

1.1 Our Contributions

Modeling the Search Engine User. Most previous work on sponsored search has (implicitly) modeled the user using two types of parameters: ad-specific click-through rates p_i and position-specific visibility factors α_j. There are some intuitive user behavior models that express overall click-through probabilities in terms of these parameters. One possibility is "for each position j *independently*, the user looks at the ad i in that position with probability α_j then clicks on the ad with probability p_i." Alternatively: "The user picks a *single* position according to the distribution implied by the α_j's, and then clicks on the ad i in that position with probability p_i." Under both these models, it follows that the probability of an ad i in position j receiving a click is equal to $p_i \alpha_j$, which is the so-called *separability* assumption [3]. From separability it follows that GSP ordering of ads will be suitable, because GSP ordering maximizes the total advertiser value on the page.

In both these models there is no reason *a priori* that the position factors α_j should decrease; this is simply imposed because it makes sense, and it is verifiable empirically. Also, both suggested models assume that the probability of an ad getting clicked is independent of *other ads* on the page, an assumption made without much justification. It is hard to imagine that seeing an ad, perhaps followed by a click, has no effect on the subsequent behavior of the user.

In designing a user model, we would like the monotonicitiy of click-through rate in position to arise naturally. Also, each ad should have parameters dictating their effect on the user both in terms of clicking on that ad, as well as looking at other ads. We propose a model based on a user who starts to scan the list of ads from the top, and makes decisions (about whether to click, continue scanning, or give up altogether) based on what he sees. More specifically, we model the user as the following Markov process: "Begin scanning the ads from the top down. When position j is reached, click on the ad i with probability p_i. Continue scanning with probability q_i." In this model, if we try to write the click probability of an ad i in position j as $p_i \alpha_j$, we get that $\alpha_j = \Pi_{i' \in A} q_{i'}$, where A is the set of ads placed above[1] position j. Thus the "position factor" in the click probability decreases with position, and does so naturally from the model. Also note that we do not have separability anymore, since α_j depends on which ads are above position j. Consequently, it can be shown that GSP assignment of ads is no longer the most efficient.

[1] Throughout the paper, we will often refer to a position or an ad being "higher" or "above" another position or ad; this means that it is earlier on the list, and is looked at first by the user.

Auction with Markovian users. Given this new user model, we can now ask what the best assignment is of ads to slots. We will study the most efficient assignment; i.e., the one that maximizes total advertiser value derived from user clicks. It turns out that the structure of this assignment is different from that of GSP, and indeed is more sophisticated than any simple ranking. The presence of the q_i's requires a tradeoff between the click probability of an ad and its effect on the slots below it. In this paper, we identify certain structural properties of the optimal assignment and use them to find such an optimal assignment efficiently, not only in polynomial time, but in near-linear time. Given this algorithm, a natural candidate for pricing is VCG [18,6,11], which is truthful in this setting under a private value model.

Monotonicity of Bidding. Now that we have defined a more sophisticated assignment function, even though VCG pricing makes the mechanism truthful under a private click-value model, the auction may not still admit the intuitive bidding strategies that are so important under GSP—especially for advertisers without a precise notion of click value. Our main technical result is to show that in our Markov model, if a mechanism uses the most efficient assignment, indeed position and click probabilities are monotonic in an ad's bid (with all other bids fixed), thus preserving this important property. Monotonicity of click probability follows from the general result of Archer and Tardos [4] on single-parameter mechanisms—for completeness we provide a proof from first principles. In contrast, position monotonicity turns out to be rather involved to prove, requiring some detailed combinatorial arguments, and insights into the optimal substructure of bidder assignments.

1.2 Related Work

Sponsored search has been an active area of research in the last several years after the early papers explored the foundational models [8,3,17,15]. In general, the motivation for the work that followed is that sponsored search in practice is much more complex than as described by the first models; see [9] for a discussion.

Only very recently are alternate user models that break the separability assumption starting to receive some attention. Ghosh and Mahdian [10] study a very general model and show hardness results for the allocation (winner determination) problem; they also give algorithms for several special cases, but none of those imply the algorithms discussed in this work. Craswell et al. [7] give an empirical study of several user click models. The "cascade" model, which was found to fit the data the best, is a special case of the model we study here (with $p_i = 1 - q_i$ and the events being mutually exclusive). Gunawardana and Meek [12] performed an empirical study of ad aggregators with the goal of detecting the affect of an ad on the other ads on the page. Their findings were consistent with our model; i.e., the presence of an ad can have a significant affect on the ads below it. Athey and Ellison [5] present a model where users have an inherent need, and click until that need is filled (or there is little chance of it getting filled). They analyze user behavior, advertiser bidding strategies and Bayesian equilibria in their model.

Independently of our work (which also appeared in [1,2]), Mahdian and Kempe [13] study the same model we do here. They also provide an $O(n \log n + nk)$ dynamic program for allocation; however at that point they generalize to the case of position-dependent continuation probabilities and provide an approximation algorithm for this case, whereas we go on to study deeper structural and incentive properties in the original model.

2 Markov User Click Model

We consider an auction with n bidders $\mathcal{B} = \{1, \ldots, n\}$ and k positions. We will also refer to "ad i," meaning the advertisement submitted by bidder i. Each bidder $i \in \mathcal{B}$ has two parameters, p_i and q_i. The click-through-rate p_i is the probability that a user will click on ad i, given that they *look* at it. The continuation probability q_i is the probability that a user will look at the next ad in a list, given that they look at ad i.

Each bidder submits a bid b_i to the auction, representing the amount that they value a click. The quantity $p_i b_i$ then represents the value of an "impression," i.e., how much they value a user looking at their ad. This is commonly referred to as their "ecpm."[2] Throughout, we use the notation $e_i = p_i b_i$ for convenience.

Our model is as follows. Given an assignment $(x_1, \ldots, x_k)$ of bidders to the k positions, the user looks at the first ad x_1, clicks on it with probability p_{x_1}, and then continues looking with probability q_{x_1}.[3] This is repeated with the second ad, etc., until the last ad is reached, or some continuation test has failed. Thus the overall expected value of the assignment to the bidders is $e_{x_1} + q_{x_1}(e_{x_2} + q_{x_2}(e_{x_3} + q_{x_3}(\ldots q_{x_{k-1}}(e_{x_k}))))$.

The goal of the auctioneer is to compute an assignment of ads to positions that maximizes overall expected value. Given this assignment, prices can be computed using VCG [18,6,11]: for each assigned bidder we compute the change in others' value if that bidder were to disappear. This assures truthful reporting of bids under a profit-maximizing utility.

3 Properties of the Optimal Assignment

In this section, we establish several properties of optimal assignments in this Markov user model, including our main technical result that position and click probability will be monotone in bid and match our intuition. We also give our algorithm for finding an optimal assignment, which gives the truthful auction via VCG pricing. All proofs can be found in [1].

[2] The acronym ecpm stands for "expected cost per thousand" impressions, where M is the roman numeral for one thousand. We will drop the factor of one thousand and refer to $p_i b_i$ as the "ecpm."

[3] The user could also have some fixed probability of looking at the first ad, which can be trivially incorporated into our results, and we leave this out for clarity. Also, the click event and the continuation event could in principle have some correlation, and all our results will still hold. However since we only consider expected value, we never use this correlation explicitly in our analysis.

Adjusted ECPM. It turns out that the quantity $e_i/(1-q_i)$, which we will refer to as the "adjusted ecpm (a-ecpm)," plays a central role in this model. Intuitively, this quantity is the impression value adjusted by the negative effect this bid has on the ads below it. We use $a_i = e_i/(1-q_i)$ for convenience. The following lemma tells us how to assign a set of k selected ads to the k positions:

Lemma 1. [4] *In the most efficient assignment, the ads that are placed are sorted in decreasing order of adjusted ecpm $a_i = e_i/(1-q_i)$.*

While this result tells us how to sort the ads selected, it does not tell us *which* k ads to select. One is tempted to say that choosing the top k ads by a-ecpm would do the trick; however the following example proves otherwise.

Example 1. Suppose we have two slots and three bidders as follows:

Bidder	e_i	q_i	$a_i = e_i/(1-q_i)$
1	\$1	.75	4
2	\$2	.2	2.5
3	\$0.85	.8	4.25

Let's consider some possible assignments and their efficiency. If we use simple ranking by ecpm e_i, we get the assignment $(2,1)$, which has efficiency $\$2 + .2(\$1) = \$2.20$. If we use simple ranking by a-ecpm a_i we get the assignment $(3,1)$ with efficiency $\$0.85 + .8(\$1) = \$1.65$. It turns out that the optimal assignment is $(1,2)$ with efficiency $\$1 + .75(\$2) = \$2.50$. The assigned bidders are ordered by a-ecpm in the assignment, but are not the top 2 bidders by a-ecpm.

Now suppose we have the same set of bidders, but now we have three slots. The optimal assignment in this case is $(3,1,2)$; note how bidder 3 goes from being unassigned to being assigned the first position.

Bidder Dominance. In classical sponsored search with simple ranking, a bidder j can dominate another bidder i by having higher ecpm; i.e., bidder j will always appear whenever i does, and in a higher position. Example 1 above shows that having a higher ecpm (or a-ecpm) does not allow a bidder to dominate another bidder in our new model. However, we show that if she has higher ecpm *and* a-ecpm, then this does suffice. This is not only interesting in its own right, it is essential for proving our deeper structural properties.

Lemma 2. *For all bidders i in an optimal assignment, if some bidder j is not in the assignment, and $a_j \geq a_i$ and $e_j \geq e_i$, then we may substitute j for i, and the assignment is no worse.*

Subset Substructure in Optimal Assignments. We show some subset structure between optimal assignments to different numbers of slots. This is used to prove position monotonicity, and is an essential ingredient of our algorithm. Let $\mathrm{OPT}(C, j)$ denote the set of all optimal solutions for filling j positions with bidders from the set C.

Theorem 1. *Let $j \in \{1, \ldots, k\}$ be some number of positions, and let C be an arbitrary set of bidders. Then, for all $S \in \mathrm{OPT}(C, j-1)$, there is some $S' \in \mathrm{OPT}(C, j)$ where $S' \supset S$.*

[4] Interestingly, this result also essentially follows from a prior work on optimizing database queries [14].

Monotonicity of Position and Click Probability. Our main theorem regarding the structure of the optimal assignments in the Markovian click model is that position and click probability are monotonic in a bidder's bid, with all other bids fixed. This is a fundamental property that makes the bidder's interaction with the system intuitive, and allows the bidder to adjust her bid intelligently without global knowledge of the other bids.

Theorem 2. *As a bidder increases her bid (keeping all other bids fixed):*

(a) the probability of her receiving a click in the optimal solution does not decrease, and
(b) her position in the optimal solution does not go down.

Computing the Optimal Assignment. A dynamic program gives an $O(n \log n + nk)$ time algorithm. for computing the optimal assignment of bidders to positions. The algorithm proceeds as follows. First, sort the ads in decreasing order of a-ecpm in time $O(n \log n)$. Then, let $F(i, j)$ be the efficiency obtained (given that you reach slot j) by filling slots $(j, \ldots, k)$ with bidders from the set $\{i, \ldots, n\}$. We get the following recurrence: $F(i, j) = \max(F(i+1, j+1)q_i + e_i, F(i+1, j))$. Solving this recurrence for $F(1, 1)$ yields the optimal assignment, and can be done in $O(nk)$ time. In fact, insights from the previous sections give an $O(n \log n + k^2 \log^2 n)$ time algorithm which is faster when k is large with respect to $\log n$:

Theorem 3. *Consider the auction with n Markovian bidders and k slots. There is an optimal assignment which can be determined in $O(n \log n + k^2 \log^2 n)$ time.*

4 Concluding Remarks

We approached sponsored search auctions as a three party process by modeling the behavior of users first and then designing suitable mechanisms to affect the game theory between the advertiser and the search engine. This formal approach shows an intricate connection between the user models and the mechanisms.

There are some interesting open issues to understand about our model and mechanism. For example, in order to implement our mechanism, the search engine needs to devise methods to estimate the parameters of our model, in particular, q_i's. This is a challenging statistical and machine learning problem. Also, we could ask how much improvement in efficiency and/or revenue is gained by using our model as opposed to VCG without using our model.

More powerful models will also be of great interest. One small extension of our model is to make the continuation probability q_i a function of location as well, which makes the optimization problem more difficult; Mahdian and Kempe [13] have given an approximation algorithm for this case, and so it is natural to ask if position monotonicity is preserved in their algorithm. One can also generalize the Markov model to handle arbitrary configurations of ads on a web page (not necessarily a search results page), or to allow various other user states (such as navigating a landing page). Finally, since page layout can be performed dynamically, one could ask what would happen if the layout of a web page were a part of the mechanism; i.e., a function of the bids.

Acknowledgments. We thank the anonymous reviews of a previous version of this paper for their comments and for pointing us to several references.

References

1. Aggarwal, G., Feldman, J., Muthukrishnan, S., Pál, M.: Sponsored search auctions with Markovian users (May 2008), http://arxiv.org/abs/0805.0766
2. Aggarwal, G., Feldman, J., Muthukrishnan, S., Pál, M.: Sponsored search auctions with Markovian users. In: Workshop on Ad Auctions (July 2008)
3. Aggarwal, G., Goel, A., Motwani, R.: Truthful auctions for pricing search keywords. In: ACM Conference on Electronic Commerce (EC) (2006)
4. Archer, A., Tardos, E.: Truthful mechanisms for one-parameter agents. In: Proc. of the 42nd IEEE Symposium on Foundations of Computer Science (2001)
5. Athey, S., Ellison, G.: Position auctions with consumer search. Levine's Bibliography 122247000000001633, UCLA Department of Economics (October 2007)
6. Clarke, E.: Multipart pricing of public goods. Public Choice 11, 17–33 (1971)
7. Craswell, N., Zoeter, O., Taylor, M., Ramsey, B.: An experimental comparison of click position-bias models. In: WSDM 2008 (2008)
8. Edelman, B., Ostrovsky, M., Schwarz, M.: Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. In: Second workshop on sponsored search auctions (2006)
9. Feldman, J., Muthukrishnan, S.: Algorithmic Methods for Sponsored Search Advertising, pp. 91–124. Springer, Heidelberg (2008)
10. Goel, G., Mehta, A.: Online budgeted matching in random input models with applications to adwords. In: SODA (2008)
11. Groves, T.: Incentives in teams. Econometrica 41(4), 617–631 (1973)
12. Gunawardana, A., Meek, C.: Aggregators and contextual effects in search ad markets. In: Workshop on Targeting and Ranking for Online Advertising (2008)
13. Kempe, D., Mahdian, M.: Cascade model for externalities in sponsored search. In: Workshop on Internet Ad Auctions (2008)
14. Krishnamurthy, R., Boral, H., Zaniolo, C.: Optimization of nonrecursive queries. In: VLDB 1986, pp. 128–137 (1986)
15. Lahie, S., Pennock, D., Saberi, A., Vohra, R.: Sponsored Search Auctions. In: Algorithmic Game Theory, pp. 699–716. Cambridge University Press, Cambridge (2007)
16. Szymanski, B.K., Lee, J.-S.: Impact of ROI on bidding and revenue in sponsored search advertisement auctions. In: 2nd Wkshp Sponsored Search Auctions (2006)
17. Varian, H.: Position auctions. International Journal of Industrial Organization 25(6), 1163–1178 (2007)
18. Vickrey, W.: Counterspeculation, auctions and competitive-sealed tenders. Finance 16(1), 8–37 (1961)