Sponsored Search Auction Design via Machine Learning*

Maria-Florina Balcan† Avrim Blum† Jason D. Hartline‡ Yishay Mansour §

ABSTRACT

In this work we use techniques from the study of sample-complexity in machine learning to reduce revenue maximizing auction problems to standard algorithmic questions. These results are particularly relevant to designing good pricing mechanisms for sponsored search. In particular we apply our results to two problems: profit maximizing combinatorial auctions, and auctions for pricing semantically related goods. Auctions for sponsored search can be viewed as combinatorial auctions in that bidders have combinatorial (in the search terms and the location of the ad on the search results page) preferences for having ads placed. Furthermore since the space of all searches is much larger than the set of advertisers, it is useful to use the semantic relationship of search terms within pricing algorithms. Our main results show how to take algorithms that solve these pricing problems and convert them into auctions with good game-theoretic properties and provably good performance.

1. INTRODUCTION

The typical approach to auctions for sponsored search is to run a separate auction for every search. This has the potential not to perform optimally as it ignores implicit competition between advertisers bidding on semantically similar keywords. This effect is more pronounced when keywords have only a few advertisers bidding on them but the semantic space of similar keywords has many advertisers. In the case where the advertisers preferences are all common knowledge, this motivates the algorithmic problem of pricing semantically related items. One of the main results of this paper is to show, when the advertisers preferences are private, how to use semantic pricing algorithms to construct an auction that takes advantage of the available semantic information.[1]

In this work, we use techniques from sample-complexity in machine learning theory to reduce the design of revenue-maximizing incentive-compatible mechanisms to algorithmic pricing questions relevant to sponsored search. When the number of agents is sufficiently large as a function of an appropriate measure of complexity of the class of solutions being compared to, this reduction produces only a $1+\epsilon$ loss in solution quality; that is, an algorithm (or β-approximation) for the standard algorithmic problem can be converted to a $(1+\epsilon)$-approximation (or $\beta(1+\epsilon)$-approximation) for the incentive-compatible design problem. We do this in a fairly general setting that includes the following as special cases:

Auction of digital goods to indistinguishable bidders. In this problem, studied in [7, 4], we have a digital good (a good of unlimited supply with zero marginal cost) and n bidders, where each bidder i has some valuation v_i between 1 and h. Our goal is to sell our good so as to make profit comparable to the best fixed price: the price p maximizing $p \times |\{i : v_i \geq p\}|$.

Attribute Auctions. Consider auctions for advertisements based on search keys. As mentioned above, a problem with having a separate auction for each key is that this might not produce enough competition to achieve good prices. Instead, we may want to group keys into categories, say having one auction for all keys related to sporting equipment, another for transportation, and so on. Given some taxonomy (or just a collection of possible groupings of keywords), we model the problem of determining the best partitioning of keywords into markets as something we call an *attribute auction.*

---

*This paper discusses results from *Mechanism Design via Machine Learning*, available as Technical report CMU-CS-05-143, as they apply to auctions for sponsored search.

†Carnegie Mellon University. {ninamf,avrim}@cs.cmu.edu
‡Microsoft Research, Mountain View, CA. hartline@microsoft.com.
§School of computer Science, Tel-Aviv University. mansour@cs.tau.ac.il. The work was done while the author was a fellow in the Institute of Advance studies, Hebrew University. This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778, by a grant no. 1079/04 from the Israel Science Foundation and an IBM faculty award. This publication reflects only the authors' views.

[1] This is a fundamentally different approach from what is known as "broad match" or "semantic match" where advertisers are automatically entered into auctions for keywords that are semantically related to their desired keyword. In particular, we will never show an advertisers ad with any keywords other than the ones they have explicitly selected.

In this problem, bidders are not indistinguishable but instead have a set of publicly-known *attributes*, such as the keywords they are interested in, and the goal is to achieve revenue comparable to the best pricing function over these attributes from some class $\mathcal{G}$. For example, [3] considers the special case of the attribute auction problem with 1-dimensional attributes and a comparison class $\mathcal{G}$ of functions that partition bidders into k contiguous "markets" and offer a separate price in each.

In the case of advertisements, $\mathcal{G}$ might correspond to partitions of keywords in the taxonomy into k categories.

Item-pricing in combinatorial auctions. Profit maximizing combinatorial auctions are another generalization of the digital good auction problem [8, 9]. In this setting we have m different items, each in unlimited supply (like a supermarket), and bidders have valuations on *subsets* of items. Our goal is to achieve revenue nearly as large as the best auction that uses item prices (assigns a separate price to each item), which is a natural comparison class. Our results imply that $\tilde{O}(mh/\epsilon^2)$ bidders are sufficient to achieve revenue close to the optimum item-pricing (assuming the algorithmic problem can be solved for the given bidders), no matter how complicated those bidders' valuations are. In fact, our bounds only require that the optimal *revenue* be large compared to mh/ϵ^2, which improves by roughly a factor of m over the results of [8].

Auctions for sponsored search can be viewed as a special case of this problem where the items on which the bidders have combinatorial preferences are the different positions that ads can be shown on the result page of a web search.

The generic type of reduction used in these settings is that given an algorithm $\mathcal{A}$ (exact or approximate) for the non-incentive-compatible optimization problem and given a set of bidders S, we will split bidders randomly into two sets S_1 and S_2, run the algorithm separately on each set (perhaps adding an additional penalty term to the objective to penalize solutions that are too "complex" according to some measure), and then apply the solution found on S_1 to S_2 and the solution found on S_2 to S_1. Sample-complexity results from machine learning theory can then give a guarantee on the quality of the results if the number of bidders is sufficiently large compared to some notion of the complexity of the comparison class or proposed solution. However, from a learning perspective, these mechanism-design settings present a number of technical challenges: in particular, the loss function is discontinuous and asymmetric, and the range of bid values may be large.

2. DEFINITIONS

We will be considering mechanism design problems of the following general form. We have a set S of n bidders, and we assume that each bidder i has some private information $priv_i$ (like how much they are willing to pay for a digital good), as well as public information pub_i (such as their location in a network). The game itself will be defined by an abstract space of legal *offers* (like an offer to sell a good at \$17) together with a mapping ρ that defines how much profit a given offer yields from a given bidder. For example, in the case of auctioning a digital good, $\rho(\text{"offer \$17"}, priv_i) = 17$ if $priv_i \geq 17$ and 0 otherwise. We can think of ρ as defining the assumption about how agents behave as a function of their private values.

DEFINITION 1. *A* **comparison class** $\mathcal{G}$ *of pricing functions is a set of functions g that map the public information of a bidder to an offer. The* **profit** *of a function g is $\sum_i \rho(g(pub_i), priv_i)$. Note that we are implicitly considering only* unlimited supply *mechanism design problems, because the profit from bidder i does not depend on whether g received profit from other bidders j.*

Given a comparison-class $\mathcal{G}$, the *algorithm design* problem is: given both the public and private information in S, find the $g \in \mathcal{G}$ of highest total profit $\text{OPT}_{\mathcal{G}}$. In our reductions, we may also want to perform "structural risk minimization", which adds additional fake penalties to different functions g based on some measure of their complexity, in which case we will need to assume we have an algorithm that optimizes revenue minus penalty. The reason for adding these penalties is that they will help to prevent the algorithm from "over-fitting" to its input: this will be important when, in our reduction, we run an algorithm on some set S_1 and apply its results to a different set of bidders S_2.

We now need to define what we mean by an incentive compatible mechanism. An incentive-compatible mechanism is a function that takes in the public information of all the bidders, plus the private information of all bidders *except* the given bidder i and outputs an offer. Our goal will be to design such a mechanism whose total profit is nearly as large as the profit of the best function in comparison class $\mathcal{G}$.

While we look to compare our profit to the profit of the best function from some class, our auction's outcome will not typically be representable as the result of using such a function. Since the auction is based on randomly partitioning the bids into two sets, the function used for each set will generally be different. This observation is not a drawback of the technique we propose nor of our performance measures.[2]

One final point at this level of generality: we will assume that we are given an upper bound h on the value of ρ; that is, no individual bidder can influence profit by more than h. This term will then come into our sample-complexity bounds.

2.1 Examples

[2] In the special case of digital-good auctions Goldberg et al. [6] give substantial justification for comparing auctions which can use multiple prices (analogously pricing functions) to an optimal single price profit: from a large class of natural auctions for profit maximization, none can beat the profit of the optimal single sale price. Furthermore, as shown by Goldberg and Hartline [5], multiple prices are inherently necessary for profit maximizing auctions: there is no truthful auction that always uses a single pricing function for all bidders and obtains an profit comparable to the optimal single price profit in worst case.

Auction of digital goods to indistinguishable bidders: As described in the introduction, in this setting the bidders have no public information (equivalently, all the bidders have the *same* public information *pub*) and the private information of bidder i is exactly its valuation v_i for the digital good, which is a real number between 1 and h. Here, a natural comparison class $\mathcal{G} = \{g_p\}$ is the class of all functions that offer a fixed price p, and ρ is a function defined by $\rho(p, priv_i) = p$ if $p \leq priv_i$ and $\rho(p, priv_i) = 0$ otherwise.

Attribute Auctions: This is the same as the setting above except now each bidder i is associated a public **attribute** $pub_i \in \mathcal{X}$ where $\mathcal{X}$ is the **attribute space**. We view $\mathcal{X}$ as an abstract space, but one can envision it as $\mathbb{R}^d$, for example. $\mathcal{G}$ is then a class of pricing functions from $\mathcal{X}$ to $\mathbb{R}_+$, such as all linear functions or all functions that partition $\mathcal{X}$ into k markets (say based on distance to k cluster centers) and offer a different price in each. The mapping ρ is a function from $\mathbb{R}_+ \times [1, h]$ to $[0, h]$ defined (as in the case of indistinguishable bidders) by $\rho(p, priv_i) = p$ if $p \leq priv_i$ and $\rho(p, priv_i) = 0$ otherwise. We will give analyses of several interesting classes of comparison functions in section 4.

Combinatorial Auctions: Here we have a set J of m distinct items, each in unlimited supply. Each consumer has a valuation $v_i(s)$ for each bundle $s \subseteq J$ of items, which measures how much receiving bundle s would be worth to the consumer i. The private information of bidder i is given by the vector of all its valuations on subsets of J (typically bidders are assumed to be indistinguishable with no public information). A natural class of comparison functions $\mathcal{G}$ (studied in [9]) is the class of functions that assign a separate price to each item, such that the price of a bundle is just the sum of the prices of the items in it (called item-pricing). The mapping ρ is then defined by assuming bidders will buy the bundle (if any) with largest positive gap between its value to them and its cost.

3. GENERIC REDUCTIONS

We are interested in reducing incentive-compatible mechanism design to the standard algorithm design problem. Our reductions will be based on Random Sampling. Let $\mathcal{A}$ be an algorithm for the (non incentive-compatible) algorithmic problem. The simplest mechanism that we consider, which we call $\text{RSOPF}_{(\mathcal{G},\mathcal{A})}$ (Random Sampling Optimal Pricing Function), is the following generalization of the random sampling digital-goods auction from [7]:

1. Randomly split the bidders into two groups S_1 and S_2, flipping a fair coin for each.
2. Run $\mathcal{A}$ to determine the best (or approximately best) function $g_1 \in \mathcal{G}$ over S_1, and similarly the best (or approximately best) $g_2 \in \mathcal{G}$ over S_2.
3. Finally, apply g_1 over S_2 and g_2 over S_1.

We will also consider variants of $\text{RSOPF}_{(\mathcal{G},\mathcal{A})}$that discretize $\mathcal{G}$ or perform some type of SRM (in which case we will need to assume $\mathcal{A}$ can optimize over the given class).

Now, fix a setting (defined by ρ and $\mathcal{G}$). In order to simplify notation, for a given pricing function g and bidder i, define $g(i)$ to be the profit made by g from bidder i, i.e., $\rho(g(pub_i), priv_i)$. Similarly, for a set of bidders $S' \subseteq S$, let $g(S') = \sum_{i \in S'} g(i)$. So, $\text{OPT}_{\mathcal{G}} = \max_{g \in \mathcal{G}} g(S)$.

The following lemma is key to our analysis.

LEMMA 1. *Consider a fixed pricing function g and a profit level p. If we randomly partition S into S_1 and S_2, then the probability that $|g(S_1) - g(S_2)| \geq \epsilon \max[g(S), p]$ is at most $2e^{-\epsilon^2 p/(2h)}$.*

We can now give our simplest generic reduction, for the case that $\mathcal{G}$ is finite. Note that for particular settings, such as the basic auction of a digital good (see [2]), we can get stronger guarantees by a more refined analysis.

THEOREM 2. *Given comparison class $\mathcal{G}$ and a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G}$, then so long as $\text{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{8h}{\epsilon^2} \ln(2|\mathcal{G}|/\delta),$$

then with probability at least $1-\delta$, the profit of $RSOPF_{(\mathcal{G},\mathcal{A})}$ is at least $(1-\epsilon)\,\text{OPT}_{\mathcal{G}}/\beta$.

In many natural cases, $\mathcal{G}$ consists of functions at different “levels of complexity” k, such as partitioning bidders into k markets. One natural approach to such a setting is to perform *structural risk minimization* (SRM), that is, to assign a penalty term to functions based on their complexity and then to run a version of $\text{RSOPF}_{(\mathcal{G},\mathcal{A})}$ in which $\mathcal{A}$ optimizes profit minus penalty. Specifically, let $\bar{\mathcal{G}}$ be a series of pricing function classes $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \ldots$, and let pen be a penalty function defined over these classes. Also for simplicity assume $\beta = 1$ (we have an exact algorithm for the underlying problem). We then define the procedure $\text{RSOPF-SRM}_{(\bar{\mathcal{G}},\mathsf{pen})}$as follows:

1. Randomly partition the bidders into two sets, S_1 and S_2, flipping fair coin for each.
2. Compute g_1 to maximize $\max_k \max_{g \in \mathcal{G}_k} [g(S_1) - \mathsf{pen}(\mathcal{G}_k)]$ and similarly compute g_2 from S_2.
3. Use price function g_1 for bidders in S_2 and g_2 for bidders in S_1.

A straightforward extension of Theorem 2 to this case would introduce a quadratic dependence in h, but we will be able to reduce this to nearly linear. Define $\text{OPT}_k = \text{OPT}_{\mathcal{G}_k}$.

THEOREM 3. *Assuming that we have an exact algorithm for solving the optimization problem required by $RSOPF\text{-}SRM_{(\bar{\mathcal{G}},\mathsf{pen})}$ then for any given value of n, ϵ, and δ, with probability at least $1-\delta$, the revenue of $RSOPF\text{-}SRM_{(\bar{\mathcal{G}},\mathsf{pen})}$ for $\mathsf{pen}(\mathcal{G}_k) = \frac{6}{(1-\epsilon)^2}\frac{72h}{\epsilon^2}\ln(8k^2|\mathcal{G}_k|/\delta)$ is*

$$\max_k \left((1-\epsilon)\,\text{OPT}_k - \mathsf{pen}(\mathcal{G}_k)\right).$$

Finally, in some cases, $|\mathcal{G}|$ is not a very good measure of the true complexity of the class $\mathcal{G}$ (e.g., even for the simplest case of fixed-price functions, if we do not discretize then $\mathcal{G}$ is infinite). In that case we can use the notion of ϵ-covers. To address this we need one more technical definition. For $g \in \mathcal{G}$ let ρ_g be the profit function induced by g and let $\rho(\mathcal{G}) = \{\rho_g : g \in \mathcal{G}\}$. That is, while g outputs an offer, ρ_g outputs the profit made from the given bidder using that offer. An ϵ-cover of $\rho(\mathcal{G})$ with respect to L_∞ is a set of functions $\mathrm{Cov}(\epsilon, \rho(\mathcal{G}))$ such that for every $\rho_g \in \rho(\mathcal{G})$ there exists f in the cover such that for every bidder i, $|\rho_g(i) - f(i)| \leq \epsilon$. Let $N(\epsilon, \rho(\mathcal{G}))$ denote the size of the smallest ϵ-cover. Now one can prove:

THEOREM 4. *If we randomly partition S into S_1 and S_2, then $n \geq \frac{8h^2}{\epsilon^2}\left(\ln\left(\frac{2}{\delta}\right) + \ln N(\epsilon/2, \rho(\mathcal{G}))\right)$ bidders are sufficient so that with probability at least $1-\delta$, for all functions $g \in \mathcal{G}$ we have $|g(S_1) - g(S_2)| \leq \epsilon n$.*

Using standard results from learning theory [1] one can bound the size of the ϵ-cover using notions such as fat-shattering dimension. However, for the special case of attribute auctions, we will get better bounds — see Section 4.2.

4. ATTRIBUTE AUCTIONS

We begin by instantiating the results in Section 3 for market pricing auctions, and then we give an analysis for general pricing functions over the attribute space that improves on the bounds of Section 3.

4.1 Market Pricing

For Attribute Auctions, one natural class of comparison functions are those that partition bidders into *markets* in some simple way and then apply a separate price in each market. For example, suppose we define $\mathcal{G}_k$ to be the set of functions that choose k bidders $b_1, \ldots, b_k$, use these as cluster centers to partition the entire set S into k markets based on distance in attribute space to the nearest center, and then offer a fixed price in each market. In that case, if we discretize prices to powers of $(1+\epsilon)$, then clearly the number of functions in $\mathcal{G}_k$ is at most $n^k (\log_{1+\epsilon} h)^k$, so Theorem 2 implies that so long as $n \geq \frac{8h}{\epsilon^2}\left[\ln(2/\delta) + k \ln n + k \ln\left(\log_{1+\epsilon} h\right)\right]$ and we can solve the algorithmic problem then with probability at least $1-\delta$, we can get profit at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}_k}$.

Another interesting and general way to do market pricing is the following. Let C be a class of subsets of $\mathcal{X}$, which we will call *feasible markets*. For k a positive integer, we consider $F_{k+1}(C)$ to be the set of all pricing functions of the following form: pick k disjoint subsets s_1,...,s_k from C, and $k+1$ prices p_0,...,p_k discretized to powers of $1+\epsilon$. Assign price p_i to bidders in s_i, and price p_0 to bidders not in any of s_1,...,s_k. For example, if $\mathcal{X} = \mathbb{R}^d$ a natural C might be the set of axis-parallel rectangles in $\mathbb{R}^d$. The specific case of $d = 1$ was studied in [3].

We can apply the results in Section 3 by using the machinery of VC-dimension to count the number of distinct such functions over any given set of bidders S. In particular, let $D = VCdim(C)$ be the VC-dimension of C and assume $D < \infty$. Define $C[S]$ to be the number of distinct subsets of S induced by C. Then, Sauer's Lemma [1] states that $C[S] \leq \left(\frac{en}{D}\right)^D$, and therefore the number of different pricing functions in $F_k(C)$ over S is at most $\left(\log_{1+\epsilon} h\right)^k \left(\frac{en}{D}\right)^{kD}$. Thus applying Theorem 2 here we get:

COROLLARY 5. *Given a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G} = F_k(C)$, then so long as $\mathrm{OPT}_\mathcal{G} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{16h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + k \ln\left(\frac{1}{\epsilon}\ln h\right) + kD \ln\left(\frac{4kh}{\epsilon^2}\right)\right],$$

then with probability at least $1-\delta$, the profit of RSOPF$_{\mathcal{G},\mathcal{A}}$ is at least $(1-\epsilon)\,\mathrm{OPT}_\mathcal{G}/\beta$.

Corollary 5 gives a guarantee in the revenue of $\mathrm{RSOPF}_{F_k(C),\mathcal{A}}$ so long as we have enough bidders n. In the following, $k \geq 0$, denote by $\mathrm{OPT}_k = \mathrm{OPT}_{F_k(C)}$. We can also show a bound that holds for all n, but with an additive loss term, as follows (we assume for simplicity here that $\beta = 1$):

THEOREM 6. *For any given value of n, k, ϵ, and δ, with probability $1-\delta$, the revenue of RSOPF$_{F_k(C),\mathcal{A}}$ is*

$$(1-\epsilon)\,\mathrm{OPT}_k - h \cdot r_F(k, D, h, \epsilon, \delta)$$

where $r_F(k, D, h, \epsilon, \delta) = O\left(\frac{kD}{\epsilon^2}\ln\left(\frac{kDh}{\epsilon\delta}\right)\right)$

Finally, we can extend our results to the setting of Structural Risk Minimization, where we want the algorithm to optimize over k, by viewing the additive loss term as a penalty function.

THEOREM 7. *Let $\bar{\mathcal{G}}$ be the sequence of pricing function classes $F_1(C), F_2(C), \ldots, F_n(C)$, and let $\mathsf{pen}(F_k(C))$ be defined appropriately. Then for any value of n with probability $1-\delta$ the revenue of RSOPF-SRM$_{\bar{\mathcal{G}},pen}$ is*

$$\max_k \left((1-\epsilon)\,\mathrm{OPT}_k - h \cdot r'_F(k, D, h, \epsilon, \delta)\right)$$

where $r'_F(k, D, h, \epsilon, \delta) = O\left(\frac{kD}{\epsilon^2}\ln\left(\frac{kDh}{\epsilon\delta}\right)\right)$.

4.2 General Pricing Functions over the Attribute Space

In this section we generalize the results in section 4.1 in two ways: to general classes of pricing functions (not just functions defined over the markets) and second, we remove the need for discretization (note that we could use results in section 3, but using the structure of the problem we show here how we can get better bounds). For example, we might want to consider a comparison class of linear functions over the attributes, or quadratic functions, or perhaps functions that divide the space into markets and are linear (rather than constant) in each market.

Assume that $\mathcal{X} \subseteq R^d$, and let $\mathcal{G}$ be a class of pricing functions over the attribute space $\mathcal{X}$. For $g \in \mathcal{G}$ let $\rho_g : \mathcal{X} \times [1, h] \to R$ be its associated profit function. Let's denote by $\rho(\mathcal{G})$ be the class of the profit functions corresponding to $\mathcal{G}$. Consider $\mathrm{OPT}_\mathcal{G} = \mathrm{OPT}(S, \mathcal{G})$ to be the profit of the optimal pricing function in $\mathcal{G}$ over S. Now, let $\mathcal{G}_d$ be the class of decision surfaces (in $\mathbb{R}^{d+1}$) induced by $\mathcal{G}$: that is, to each $g \in \mathcal{G}$ we associate the set of all $(x, v) \in \mathcal{X} \times [1, h]$ such that $g(x) \leq v$. Finally, let $D = VCdim(\mathcal{G}_d)$. Assume in the following that $D < \infty$. Then we can prove that ([2]):

THEOREM 8. *Given class $\mathcal{G}$ and a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G}$, then so long as $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{64h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + D\ln\left(\frac{64h}{\epsilon^2}\left(\frac{16}{\epsilon}\ln h + 1\right)\right)\right],$$

then with probability at least $1-\delta$, the profit of $RSOPF_{(\mathcal{G},\mathcal{A})}$ is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$.

5. COMBINATORIAL AUCTIONS

For the case of combinatorial auctions described in Section 2.1, where we want to achieve revenue nearly as high as the best set of item-prices, we can directly apply Theorem 2. Specifically, let $\mathcal{G}$ be the class of item prices, discretized to powers of $(1+\epsilon)$. Then we have:

COROLLARY 9. *Given a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G}$, then so long as $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{8h}{\epsilon^2}\left(m\ln(\log_{1+\epsilon} h) + \ln(2/\delta)\right),$$

then with probability at least $1-\delta$, the profit of $RSOPF_{\mathcal{G},\mathcal{A}}$ is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$.

Auctions for sponsored search are combinatorial in nature. Often several advertisements are shown with the outcome of a search and advertisers may have a preference over the relative position of their ad. Furthermore, an advertiser might also have their ad shown on searches for several different keywords and may have a preference over the keywords. Item pricing is natural for these settings and the results above apply.

6. CONCLUSIONS

In this work we have made the connection between machine learning and mechanism design explicit. In doing so, we obtain a unified approach to considering a variety of profit maximizing mechanism design problems including many that have been previously considered in the literature. These results are particularly relevant to designing good pricing mechanisms for sponsored search.

7. REFERENCES

[1] M. Anthony and P. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[2] M.-F. Balcan, A. Blum, J. Hartline, and Y. Mansour. Mechanism design via machine learning. 2005. Technical Report, CMU-CS-05-143.

[3] A. Blum and J. Hartline. Near-Optimal Online Auctions. In *Proc. 16th Symp. on Discrete Alg.* ACM/SIAM, 2005.

[4] A. Fiat, A. Goldberg, J. Hartline, and A. Karlin. Competitive Generalized Auctions. In *Proc. 34th ACM Symposium on the Theory of Computing*. ACM Press, New York, 2002.

[5] A. Goldberg and J. Hartline. Envy-Free Auction for Digital Goods. In *Proc. of 4th ACM Conference on Electronic Commerce*. ACM Press, New York, 2003.

[6] A. Goldberg, J. Hartline, A. Karlin, M. Saks, and A. Wright. Competitive auctions and digital goods. *Games and Economic Behavior*, 2002. Submitted for publication. An earlier version available as InterTrust Technical Report STAR-TR-99.09.01.

[7] A. Goldberg, J. Hartline, and A. Wright. Competitive Auctions and Digital Goods. In *Proc. 12th Symp. on Discrete Algorithms*, pages 735–744. ACM/SIAM, 2001.

[8] Jason Hartline and Andrew Goldberg. Competitive auctions for multiple digital goods. In *ESA*, 2001.

[9] V. Guruswami and J. Hartline and A. Karlin and D. Kempe and C. Kenyon, and F. McSherry. On Profit-Maximizing Envy-Free Pricing. In *Proc. 16th Symp. on Discrete Alg.* ACM/SIAM, 2005.