Using Unsuccessful Auction Bids to Identify Latent Demand

Bernardo A. Huberman, Tad Hogg and Arun Swami
Xerox Palo Alto Research Center
Palo Alto, CA 94304

December 11, 2000

Abstract

We propose using the information revealed through auctions, including in particular the unsuccessful bids, to identify latent demand. Applied to combinatorial auctions for bundles of goods, this information can identify new bundles with particularly high valuations, expressed by their high complementarity. We present a simple algorithm for identifying these bundles, suitable for use with agent-based ecommerce systems.

1 Introduction

The emergence of a global electronic economy with universal access, as mediated by the Internet, makes it possible to have access to information about the behavior of individuals on a staggering scale. While most of the data generated by this economy has so far been used by businesses for marketing purposes, there are aspects of this bounty that allow for more subtle exploitation. These aspects appear as a kind of informational side effect, in which apparently uninteresting data generated by users can be put to good use. These side effects were previously unavailable, but they are now easily captured by the digital infrastructures of the electronic economy. An interesting example of this novel approach is provided by technologies that exploit the fact that mobile phones are always locating their transmission cells to make inferences about the amount of car traffic in particular sections of a city [4].

In this paper we uncover and exploit a novel informational side effect of particular relevance to electronic-commerce applications; the extra information on demand (or supply) for bundles of goods revealed by combinatorial auctions [1].

Auctions provide a market mechanism for efficiently matching buyers and sellers of many products and services. A wide variety of auction mechanisms exist with various suitabilities for different situations [5, 13, 14, 18]. Moreover, the development of electronic commerce provides many new opportunities for auctions involving automated bidding agents [19].

Auctions can be viewed as making the participants reveal some information about their preferences (e.g., a price they are willing to pay) to an auctioneer. After collecting these preferences, the auction mechanism determines which parties actually exchange the goods and at what price.

This auction mechanism typically makes use of only a very limited portion of the submitted information, e.g., the highest bid, while discarding the remainder. However, the remaining information, particularly the unsuccessful bids, provides a unique insight into

[1] U.S. Patent pending.

	bundles						
	A	B	C	AB	AC	BC	ABC
agent 1	1	2	**6**	3	9	6	10
agent 2	2	1	4	**12**	8	7	12
agent 3	1	2	3	3	7	6	10
agent 4	3	1	2	4	7	4	9
average	1.75	1.5	3.75	5.5	7.75	5.75	10.25
per item	1.75	1.5	3.75	2.75	**3.875**	2.875	3.417

Table 1: Valuations of 4 agents for combinations of 3 items, A, B and C. The boxes show the highest-value allocation. The bottom two rows show the average valuation for each bundle and the average value per item.

the potential market for the goods involved or similar ones. In particular, these bids give some indication of the demand (from buyers) or supply (from sellers) curves in the market. When aggregated over many similar auctions, such information can be useful in identifying potential markets (e.g., estimates of likely purchase volume at different prices) thereby enabling more effective pricing policies. Importantly, since the preferences represent values the participants would actually pay or accept (i.e, their own money is on the line) they should be more accurate reflections of preferences than simple polls or survey groups.

This idea becomes even more powerful in the case of combinatorial auctions, where instead of a single good being traded, the auction involves arbitrary combinations of a set of goods [1, 12]. Typically, such auctions are appropriate when buyers or sellers value a set of goods differently from the sum of their individual values. In such auctions, participants submit preferences regarding various sets of the goods. Possible applications include allocating resources on the space shuttle and determining useful combinations of items to cache or warehouse. In general, finding the winning combinations in such auctions is NP-complete, leading to the development of a variety of effective algorithms or tractable restricted versions [20, 17, 23].

While the resulting auction mechanism is more complex than the single-good case, combinatorial auctions can result in far more information from the unsuccessful bids, namely identifying combinations of the goods that are particularly highly valued. Because of the combinatorics involved, attempting to survey the market for all possible combinations of a large set of goods would be prohibitive. Thus using the information revealed by the unsuccessful bids in a combinatorial auction is not only likely to be more accurate but also more computationally efficient in that the participants themselves identify the sets of goods of particular interest to them, thereby avoiding the need to examine all possibilities.

In what follows, we describe how this additional information can also be used to estimate demand curves, thereby exposing potentially useful new product bundles. We focus on identifying potential buyer demand, though the techniques can also apply to the asking prices of sellers to estimate supply curves. The techniques we introduce for using information side-effects of auctions are particularly well-suited for simple software agents acting on the behalf of users.

2 Examples of Complementarity

To illustrate these ideas, consider the complete valuations shown in Table 1. In this case, the highest-value allocation is item C to agent 1 and items AB to agent 2, as would be determined through a generalized Vickrey auction [24]. Notice however the relatively high complementarity assigned to the combination AC which the agents value considerably more than the sum of the two items A and C individually. Even though AC is not part of the winning allocation, this information indicates the existence of a complementarity that can be potentially exploited by offering this particular bundle.

These observations are quantified by the average value per item for each bundle, shown in the last line of the table. Assuming a competitive market, so prices are not influenced by choices of individual suppliers, and unit cost for producing each item, this value indicates the potential profit from supplying the various bundles. More generally, we would include any variations in the individual item costs and complementarities in their production to determine the ratio of average values, given in the penultimate line of the table, to the corresponding costs.

3 Estimating Demand

We consider demand for bundles of n goods. As a specific model of valuations [2], suppose a person has a value v_i for the i-th good, and the value for a set S of these goods is the sum of the individual values raised to the power α. A value $\alpha = 1$ indicates no particular value for bundles beyond their individual worth, while $\alpha > 1$ indicates the bundle is worth more than the sum of its constituents, i.e., the goods are *complementary*. Alternatively, a value of $\alpha < 1$ indicates that the constituents are substitutes.

Since people can vary in their valuations of the goods and possible bundles, we would like to estimate a typical or average value of α. The simplest case assumes a constant α to be estimated from the available bids.

More generally, particular bundles may be valued more highly than others, so α could vary from bundle to bundle. Specifically, α is defined in terms of the value of a set S in terms of its components as

$$\text{value}(S) = \left(\sum_{i \in S} v_i \right)^{\alpha} \tag{1}$$

where v_i is the value associated with the individual item i by this agent. In general, the degree of complementarity will depend on the set of items S.

One use for such an estimate is for suppliers to identify bundles that are likely to be particularly profitable, i.e., with large differences between the inferred demand and the cost to produce the bundle. Such bundles could then be offered for sale. A further consequence, that may be more significant over time, is revealing the excess demand for bundles and thereby suggesting complementary uses that the suppliers may not even be aware of. Such knowledge could be useful in developing new products or business models. Significantly, in the context of software agents attempting to automatically identify potentially useful transactions, this evaluation is computationally simple in not requiring detailed reasoning about properties of various bundles.

4 Identifying High-Value Bundles

The example shown in Table 1 assumed that information was available on all possible combinations of items. In general, agents may not bid on all bundles. Given that the number of bundles grows exponentially with the number of items, incomplete information is the norm in most auctions. What is available is a *sample* from the set of all possible bids. If the number of items is large, some high value bundles may not have any registered bids in the sample. We are interested in discovering such high value bundles and assigning some likely values to these bundles.

If the form of complementarity is known, e.g., the model of Eq. 1 , precise value estimates could be made. However, this level of detailed knowledge is unlikely to be available. In what follows, we describe an algorithm for estimating bundle values systematically without making strong assumptions on the precise form of complementary valuations.

Specifically, we suppose the individual item values are known (e.g., from their prices in existing single-item markets) and we are given a sample of bids for m sets of two or more items: $S_1, \dots, S_m$. In general, the sample could include duplicate sets, e.g., from bids from different agents on the same set of items.

For each set in the sample, Eq. 1 defines a corresponding complementarity factor α_m. Based on this sample we define estimated α values for any set S as follows:

- if S appears (one or more times) in the sample, we estimate α as the average of the α values associated with the bids for S in the sample
- if neither S nor any subset of S appears in the sample, we take $\alpha = 1$ for S, i.e., assume it has no complementarity
- otherwise, the estimate for α for set S is derived from a weighted average of the α values associated with its subsets appearing in the sample, with larger subsets given more weight. Furthermore, we include a dilution factor that gives less contributions from small subsets in the sample even if they are all the same size so they would otherwise be weighted equally. Specifically, the estimate is

$$\alpha(S) = 1 + \left\langle (\alpha_i - 1) \frac{|S_i| + 1}{|S|} \right\rangle \tag{2}$$

  where $\langle ... \rangle$ denotes the weighted average over the members of the sample S_i that are subsets of S and $|S|$ is the number of items in set S.

There are a number of choices for the weightings used in the average of Eq. 2 that give more weight to larger subsets. For definiteness, we make the simple choice of weighting according to $1/d(S, S_i)$ where d is the distance between the sets which, since S_i is a subset of S, is $|S| - |S_i|$, i.e., the number of items in S that do not appear in the subset S_i, and is greater than or equal to 1. That is we take the weighted average of a function f, mapping sets to numbers, to be

$$\langle f(S) \rangle = \frac{\sum \frac{f(S_i)}{d(S,S_i)}}{\sum \frac{1}{d(S,S_i)}} \tag{3}$$

with the sums over those sets S_i in the sample that are subsets of S. Thus, when subsets in the sample have different sizes, the α values associated with the larger ones are weighted more heavily. As described above, the weighted average is used only when S itself does not appear in the sample, so $d(S, S_i) > 0$.

The factor $\frac{|S_i|+1}{|S|}$ in Eq. 2 gives full value to complementarity from subsets with one less item than S and lower values when S_i is much smaller than S, capturing the intuition that complementarity estimates are likely to be less reliable when there is a large difference between S and S_i. For example, when estimating the complementarity of a set with four items, a sample consisting of two-item sets is likely to be less informative than a sample with three-item sets. Introducing this factor in the estimation of complementarity is a simple way to bias consideration toward sets only a bit larger than those in the sample. More generally, one could capture this intuition on the estimate reliability by separating the valuation estimate from its reliability and then using porfolio approaches [11, 6] to find appropriate balances between bundles with high estimated complementarities and those with reliable estimates. This more sophisticated approach would become more appropriate when extensive data on valuations is available for particular groups of products, allowing reasonable estimates of valuations and their variances.

Although this algorithm need not capture precisely the valuations, even on average, it's simple form allows agents to rapidly sift through large amounts of data to suggest possible complementary combinations. After estimating α for the sets based on the sample, we can quantify the potential profit from supplying each bundle. If, for simplicity, we suppose all items have unit cost to create and provide no complementarity on the supply side, the cost for a bundle is just proportional to the number of items. In this case, a suitable criterion is the estimated value of a bundle divided by the number of items. More generally, if costs vary among the items and sets, we can divide the estimated value by the actual cost. Moreover, although we have discussed estimating values for buyers, the same approach could be used to estimate costs for sellers who participate in declining-price or double auctions.

In practice, this estimation algorithm can be used in a number of ways. The simplest approach, appropriate when the number of items is not too large, is to estimate the values for all sets of items and select the set with the highest estimated value to cost ratio.

As another approach, we could restrict attention to sets just somewhat larger than those in the sample, avoiding the combinatorics of considering all sets. This is particularly appropriate since the approximation of Eq. 2 for the complementarities of sets is likely to be more reliable for sets close to those in the sample than for larger sets.

Finally, we could examine sets in order of increasing size and use an upper bound on Eq. 2 to determine when the estimated value to cost ratio of larger sets will not exceed the largest value already found. At this point there would be no need to continue with additional set evaluations to find the one with the highest estimated ratio. For example, if k is the size of the largest set in the sample and $\alpha_{\max}$ is the largest α value for the sample sets, then Eq. 2 gives α for a set of size h as no larger than $\alpha_{\text{bound}}(h) = 1 + (\alpha_{\max} - 1)(k+1)/h$. Defining V_h as the sum of the h largest individual values, with Eq. 1 the estimated value for set S would be at most $V_h^{\alpha_{\text{bound}}(h)}$. This could be used as a bound to terminate the algorithm without evaluating larger sets.

5 Algorithm Behavior

To quantify the behavior of this algorithm, ideally one would examine data from real auctions in which the true valuations are independently known. Unfortunately such complete information is unlikely to be available in auctions with many items due to the large number of possible sets and the difficulty of obtaining the true valuations from participants. This limitation could be addressed through laboratory experiments [16, 21] where, through suitable experimental protocols, the valuations are specified a priori. A simpler approach, which we adopt here, uses agent-based economic simulations [8, 19]. While less realistic than studies of actual human behavior in markets, they allow examining larger cases and using many samples. Agent-based studies are also particularly appropriate when simple agents are likely to be actively used in e-commerce transactions, e.g., due to the large number of possibilities to consider in a combinatorial auction. Thus in this section we examine the behavior of the algorithm when valuations are determined by a simple functional form.

We consider a simple model with minimal assumptions on how bundles derive their values from smaller bundles. Our use of valuations generated from simple models is similar to prior studies of the performance of algorithms identifying winning combinations from bids in a combinatorial auction [20, 17]. However, instead of random variation in valuations among bundles, we focus on a situation with considerable complementarity, for particular bundles that do not appear explicitly in the bids.

Specifically, suppose one set of items, G, is particularly complementary in value for most

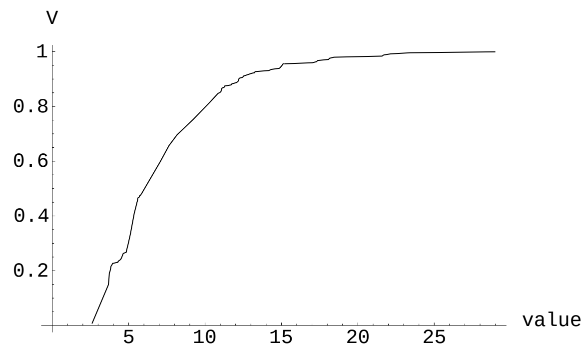

Figure 1: Cumulative distribution of values for the 247 sets of two or more bundles for the example described in the text.

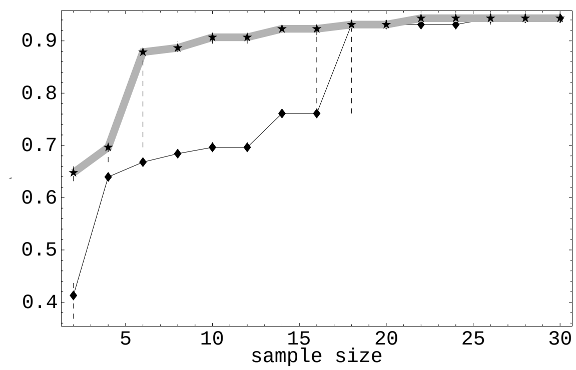

Figure 2: Median location in the cumulative value distribution of the best bundle returned by the algorithm (gray) compared with the best in the sample (black) as a function of the number of sets m in the sample. This example involves $n = 8$ items and $G = \{1, 2, 3, 4\}$ while the sets in the samples are uniformly selected among sets of size 2 and 3. The error bars show the 95% confidence intervals for the medians [22].

of the agents. In this case, as a simple valuation model, the value of α in Eq. 1 for set S can be taken to be equal to

$$\alpha = 1 + \frac{|S \cap G|}{|S \cup G|} \tag{4}$$

Thus as S includes more items of G, the complementarity increases but this is counterbalanced when S includes many additional items not found in G, corresponding to the dilution in value brought about by additional items. Furthermore, this model gives a maximum α when $S = G$.

As an example, we take the individual item values to be generated from a lognormal distribution [3] with parameters 1 and 0.2, i.e., generated by taking the exponential of values generated according to a normal distribution with mean 1 and standard deviation 0.2. Thus each item has a positive value and the distribution of individual values has a somewhat extended tail. An example for $n = 8$ items and $|G| = 4$ produces the cumulative distribution of values shown in Fig. 1 .

Fig. 2 shows the ability of the algorithm described above to identify high-value sets based on a sample of valuations. Note that the algorithm does not make use of the particular functional form for the complementarity used in this example (i.e., Eq. 4), but rather the general property of complementarity that a high-value set is likely to have relatively high-value subsets, on average.

We see that the algorithm identifies higher percentile sets based on small samples than simply picking the best set in the sample. We also found similar behavior with other choices for the number of items, sample sizes and individual values.

6 Improving the Estimate

Since the number of possible bundles grows very rapidly with the number of items, in general one cannot expect to have enough information from actual auctions to fully determine the demand for different bundles. Moreover, the value of different bundles may not follow the simple model of complementary costs assumed above, i.e., the complementarity of a set is similar to that of its subsets, on average.

This difficulty can be approached in a number of ways. First, one could focus on bundles with relatively few items, thereby reducing the number of possibilities to consider. However, such small bundles are also the most likely to have already been noticed and thereby offer limited gains for use with this technique. A second approach assumes a degree of uniformity of demand with respect to possible bundles so that the limited available information is sufficient to extrapolate over most of the sets. A third possibility involves obtaining specific additional information that best helps resolve ambiguities to distinguish particularly highly-valued bundles, even if they involve a fairly large number of items.

One way to increase the information available is to perform a few additional auctions that particularly test for bundles not otherwise sampled in the existing data. To encourage participation, these additional auctions may require subsidies, in effect purchasing the information as a type of information market [7]. Since these will incur additional overhead, either in time or money, compared to using the existing auction results, it is important to extract as much information as possible about the bundle valuations with just a few additional "probes".

This strategy raises the question of how best to sample the possible sets most efficiently when we can specifically choose additional samples. On the one hand, it could be useful to sample small sets of items that did not appear among the previous auctions, i.e., scattering the sample points as far apart as possible. On the other hand, particular complementarities may only be revealed in somewhat larger sets, even though they include many sets already examined in previous auctions. Thus we require how best to sample not only broadly among sets of a particular size but also sets of different sizes. One approach is to focus on adding samples to maximally increase the diversity of sets examined. In hierarchical structures, diversity can be quantified by the number of distinct substructures [9]. Generalizing to other combinatorial structures, such as sets, may prove useful in identifying a few additional samples with the most impact on improving valuation estimates. Such an approach, though comute intensive, is particularly appropriate for agents with limited reasoning sophistication and lacking detailed knowledge of item uses.

7 Discussion

Making use of the additional information from auctions for an aggregate population gives a general idea of new market opportunities. A more specific targeting is possible if the information is associated with individual bidders. However, using such information raises two issues for designing the mechanism. First is incentive compatibility: if users are aware that their bids will be used not only for the auction at hand but also for targeted marketing, they may benefit from misrepresenting their preferences (e.g., if they believe they may be offered the good by another company at the price of their bid submitted to the auction). Such misrepresentation not only reduces the accuracy of the collected information, but could also be detrimental to the underlying auction. Since incentive issues already arise in designing combinatorial auctions themselves [25], it is important to keep this issue in mind when adding new uses for the bid information.

A second issue is a more general concern over privacy, e.g., participants may be reluctant to even enter a particular auction if their interest in it will be made available to others. Such concerns can be addressed by anonymous participants, and can still allow aggregate identification of potential markets. However, this looses the ability to provide individual pricing based on past behavior (e.g, additional discounts to a customer who has purchased a lot in the past). Existing cryptographic techniques for secure distributed computation provide an alternative in which the information made available can be precisely limited [10, 15] to the minimum required for a given type of transaction. Such techniques enable a wide-range of trade-offs between privacy concerns and increased economic efficiency from using a broader range of information about market participants. They could be especially useful in the context of agents acting on behalf of people by preventing inadvertent information disclosure.

The growth of electronically mediated auctions makes it feasible to aggregate and use behavior revealed in the auctions for identifying demand. Furthermore, with the development of software agents to assist users in finding suitable transactions and even participating directly in auctions on behalf of users, the use of electronic auctions could expand to a much wider range of transactions than are used today. Thus we can expect a growing source of auction information, thereby improving the demand estimates made possible by using the unsuccessful bids. More broadly, expanded electronic mediation of transactions will provide numerous opportunities to utilize information side-effects from these transactions.

Acknowldgements

This work was partially supported by a National Science Foundation Grant No. NSF IIS-9986651.

References

[1] Christopher Avery and Terrence Hendershott. Bundling and optimal auctions of multiple products. *Review of Economic Studies*, 67:483–497, 2000.

[2] Yannis Bakos and Erik Brynjolfsson. Bundling information goods: Pricing, profits and efficiency. *Management Science*, Dec. 1999.

[3] Edwin L. Crow and Kunio Shimizu, editors. *Lognormal Distributions: Theory and Applications*. Marcel Dekker, New York, 1988.

[4] Sam Diaz. Cell phone signals touted to fight traffic wars. *Silicon Valley News*, Jan. 20 2000. Available at www0.mercurycenter.com/svtech/news/indepth/docs/-traf012100.htm.

[5] Daniel Friedman. The double auction market institution: A survey. In D. Friedman and J. Rust, editors, *The Double Auction Market Institutions, Theories, and Evidence*, pages 3–25, NY, 1993. Addison-Wesley.

[6] C. P. Gomes and B. Selman. Algorithm portfolio design: Theory vs. practice. In D. Geiger and P. Shenoy, editors, *Proc. of the 13th Conf. on Uncertainty in AI (UAI-97)*, pages 190–197, Los Altos, CA, 1997. Morgan Kaufmann.

[7] Robin Hanson. Decision markets. *IEEE Intelligent Systems*, pages 16–19, May/June 1999.

[8] John H. Holland. Artificial adaptive agents in economic theory. Technical Report 91-05-025, Santa Fe Institute, Santa Fe, NM, 1991.

[9] B. A. Huberman and T. Hogg. Complexity and adaptation. *Physica*, 22D:376–384, 1986.

[10] Bernardo A. Huberman, Matt Franklin, and Tad Hogg. Enhancing privacy and trust in electronic communities. In *Proc. of the ACM Conference on Electronic Commerce (EC99)*, pages 78–86, NY, 1999. ACM Press.

[11] Bernardo A. Huberman, Rajan M. Lukose, and Tad Hogg. An economics approach to hard computational problems. *Science*, 275:51–54, 1997.

[12] Paul Klemperer. What really matters in auction design. Technical report, Oxford University, June 2000.

[13] R. Preston McAfee and John McMillan. Auctions and bidding. *J. of Economic Literature*, XXV:699–738, June 1987.

[14] Kevin A. McCabe, Stephen J. Rassenti, and Vernon L. Smith. Smart computer-assisted markets. *Science*, 254:534–538, October 25 1991.

[15] Moni Naor, Benny Pinkas, and Reuben Sumner. Privacy perserving auctions and mechanism design. In *Proc. of the ACM Conference on Electronic Commerce*, pages 129–139, NY, 1999. ACM Press.

[16] Markus Noeth, Colin F. Camerer, Charles R. Plott, and Martin Weber. Information aggregation in experimental asset markets: Traps and misaligned beliefs. Social Science Working Paper 1060, Caltech, April 1999.

[17] David C. Parkes and Lyle H. Ungar. Iterative combinatorial auctions: Theory and practice. In *Proc. of the 17th Natl. Conf. on Artificial Intelligence (AAAI2000)*, pages 74–81. AAAI, 2000.

[18] Charles R. Plott. Laboratory experiments in economics: The implications of posted-price institutions. *Science*, 232:732–738, May 9 1986.

[19] John Rust, Richard Palmer, and John H. Miller. Behavior of trading automata in a computerized double auction market. In D. Friedman, J. Geanakoplos, D. Lane, and J. Rust, editors, *The Double Auction Market: Institutions, Theories and Evidence*. Addison Wesley, 1992.

[20] Tuomas Sandholm. An algorithm for optimal winner determination in combinatorial auctions. In *Proc. of Intl. Joint Conf. on AI (IJCAI99)*, pages 542–547, 1999.

[21] Vernon L. Smith. Experimental methods in the political economy of exchange. *Science*, 234:167–234, 1986.

[22] George W. Snedecor and William G. Cochran. *Statistical Methods*. Iowa State Univ. Press, Ames, Iowa, 6th edition, 1967.

[23] Moshe Tennenholtz. Some tractable combinatorial auctions. In *Proc. of the 17th Natl. Conf. on Artificial Intelligence (AAAI2000)*, pages 98–103. AAAI, 2000.

[24] H. Varian and J. K. MacKie-Mason. Generalized Vickrey auctions. Technical report, Univ. of Michigan, 1995.

[25] Makoto Yokoo, Yuko Sakurai, and Shigeo Matsubara. Robust combinatorial auction protocol against false-name bids. In *Proc. of the 17th Natl. Conf. on Artificial Intelligence (AAAI2000)*, pages 110–115. AAAI, 2000.