Modeling the value of information granularity in targeted advertising

Modeling the value of information granularity in targeted advertising

Swati Rallapalli[†], Qiang Ma[‡], Han Hee Song[§], Mario Baldi[§], S. Muthukrishnan[‡], Lili Qiu[†]
The University of Texas at Austin[†] Rutgers University[‡] Narus Inc.[§]

ABSTRACT

Behavioral Targeting (BT) in the past few years has seen a great upsurge in commercial as well as research interest. To make advertising campaigns more effective, advertisers look to target more relevant users. Ad-networks and other data collectors, such as, Cellular Service Providers (CSPs), hold a treasure trove of user information that is extremely valuable to advertisers. Moreover, these players may have complimentary sets of data. Combining and using data from different collectors can be very useful for advertising. However, in the trade of data among the various players, it is currently unclear how a price can be attached to a certain piece of information. This work contributes (*i*) a MOdel of the Value of INformation Granularity (*MoVInG*) that captures the impact of additional information on the revenue from targeted ads in case of uniform bidding and (*ii*) an expression that is applicable in more general scenarios. We apply MoVInG to a user data-set from a large CSP to evaluate the financial benefit of precise user data.

1. INTRODUCTION

The budget invested in Behavioral Targeting (BT) in the year 2012 is estimated to be $4.4 billion [6]. In the year 2009, 24% of online marketers used BT [7] a 50% increase from the previous year. Studies show that the conversion rates (*i.e.*, fraction of people who buy an item/service after seeing its ad) of targeted advertisements are twice that of the non-targeted ones [5].

In order to make targeting more effective for advertisers, ad-networks not only collect user data, *e.g.*, browsing behavior, demographics, etc., but also trade the user-data with each other (while adhering to their privacy terms and conditions) [8, 14]. However, it is unclear what price tag can be attached to a certain piece of information. For example, what is the impact of finer-granularity user-data on the conversion rate and consequently on the revenue that ad-networks might expect from targeted ads? Consider, for example, a car dealer in San Francisco, California, who wants to advertise a special offer available to people living in her city. A data-set with coarse-granularity, *e.g.*, with residence information at the state level, is of less use compared to a data-set that has city level information. In fact, ads placed using the state-level data-set will likely reach many people not interested in the offer. Consequently, the conversion rate for ads placed based on the finer-grained data-set will likely be higher and thus such a data-set is more valuable.

Currently, most of the user profiles are collected by partner networks of the web services that the users visit. By combining the history of the user footprints at each partner service, the players in the ad industry form a picture of the user's profile. This *service-side* profiling method lacks ways to combine information from multiple devices, multiple browsers, and web services that are not partners in profile collection. There are other players, *e.g.*, ISPs, CSPs, who can build rich and fine-grained user data-sets by profiling their users. This *network-side* profiling can be extremely valuable for targeting because as long as a user is on the same edge ISP (or CSP), her data can be collected regardless of her devices, OSes, services, etc. However, in order for these players to take full advantage of the benefit their data can offer, they must be able to properly price it. This work devises a MOdel of the Value of INformation Granularity (*MoVInG*) by deriving expressions to quantify the impact of such fine-grained data on the revenue from advertising.

Challenges: Estimating the benefit of fine-grained user data is challenging due to the following reasons: (1) lack of measures for data *quality* that take into account different granularities for various attributes; (2) several factors affecting *competition* among advertisers, *e.g.*, different advertisers are interested in placing ads on overlapping sets of users based on different attributes; (3) several *proprietary components*, *e.g.*, targeting algorithm. MoVInG does not relate the impact of user data on the revenue to information granularity explicitly, which can be extremely challenging given the variety of aggregation levels of different attributes. Instead, MoVInG derives the relationship between revenue from user data and the number of redundant users that might be included in a target set due to the lack of fine-grained information.

Contributions: (*i*) We derive MoVInG to capture the difference in revenue resulting from different data-sets, considering a pay-per-impression payment model where advertisers pay based on the number of ads (impressions) shown (Section 2.1). We assume a single-sided auction market for ad slots, where advertisers draw bids from a uniform distribution. (*ii*) We validate the obtained model by simulating the auction under both the assumption of uniform bid distribution and more general bidding behaviors. Since the information granularity value as computed by MoVInG refers to a specific bidding scenario, (*iii*) we present a way of deriving a more general expression, based on few input parameters and we show that it fits the simulation results very well (Section 2.2). Finally, (*iv*) we present concrete examples of the gain in revenue due to improved user-data precision using the traffic trace of a large North-American CSP (Section 3).

2. REVENUE MODEL

Let D be a more precise data-set that has finer-grained information than a less precise D'. To quantify the impact of improved precision, we want to derive the difference in the revenue of the ad-network when ad targeting is based on D versus D'.

Figure 1 shows a schematic view of the players in the targeted ad-

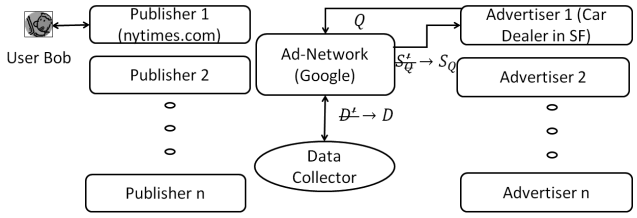


Figure 1: Advertisement placement architecture.

Notation	Details
D	Precise data-set
D'	Imprecise data-set
A	Number of advertisers
U	Total number of users in the system (in D and D')
Q	Advertisers' query on user attributes in the data-set
$S_Q(S'_Q)$	User-set returned by query on D, $Q(D)$ (D', $Q(D')$)
$S(S')$	Expected size of S_Q (S'_Q) over all queries
$c(c')$	Cost per impression/click when using D (D')
a_i	Attribute i of user information
v_i	Value for attribute i selected by query Q
L	Loss of revenue (between c and c')
$n(n')$	Number of advertisers competing per slot in D (D')
r	Maximum possible bid of an advertiser

Table 1: Notations.

vertising architecture. *Publishers* own websites (*e.g.*, nytimes.com), generate the content for them, and reserve ad spaces termed as *slots*. *Advertisers* (*e.g.*, car dealer in San Francisco) are interested in getting the attention of the viewers of the web pages towards their items/services by displaying ads. *Ad-Networks* (*e.g.*, Google AdSense) are intermediaries deciding which ad is shown on which publisher's page. We consider the following ad placement architecture: (1) advertisers specify the user population they are interested in by means of a query Q (*i.e.*, they are interested in users satisfying Q). Q specifies the values that selected user attributes must have in order for the advertiser to be interested in a user; (2) the ad-network returns a list of users S_Q or S'_Q who best match the query $Q(D)$ or $Q(D')$, depending on whether the ad-network by itself or along with a partner data collector owns D or D' and (3) the advertisers, in turn, bid on the web page visits by any of the S_Q or S'_Q users, and the winning advertiser's ad is placed on the publisher's pages. Note that our findings are general and are applicable in other ad-placement scenarios *e.g.*, where ad-network carries out the ad-placement directly on the pages visited by the users.

Next we define the terms used. *Valuation* is the amount advertisers consider an ad is worth and, consequently, it determines their bids. *Bid* is the maximum amount an advertiser is willing to pay to display an ad. *Revenue* is the gross income of an ad-network resulting from targeted advertising before any expense deductions and the cut to the publishers. Table 1 summarizes our notations.

2.1 Information Value Model

We identify two factors that affect ad-networks' revenue: (*i*) valuation of the advertisers for the impression, which goes *lower* with imprecise data-set, as explained in Section 1; and (*ii*) amount of competition during the auction of each ad slot, which goes *higher* with imprecise data-set. For example, consider the user sets returned to one car dealer in San Francisco and another in Los Angeles, both interested in local customers. They are disjoint when based on data-set D (which has city level information for user location), hence advertisers do not compete for any slot. However, with D' (which has only state level information), both advertisers are returned a set of all users in California and consequently they compete in advertising to common users. This *artificial competition* actually improves ad-networks' revenue.

In the derivation, we adopt an incremental approach and at first consider an advertisement placement that is not based on auctions and next consider ad auction scenarios.

Fixed Price Advertisements: Let's consider a fixed cost c and c' that an advertiser pays for target users based on data-set D and D', respectively. When using D', ads are displayed to users who are actually not in the target audience because $S_Q \in S'_Q$. Hence impression cost $c' < c$. We assume $c' = c * \frac{S}{S'}$, the intuition being that the advertiser will be willing to pay only for the fraction of ads shown to the real target audience, i.e., the fraction of the users in S'_Q that also belong to S_Q. Hence loss of revenue per impression:

$$L = c - c' = c(1 - \frac{S}{S'}) \quad (1)$$

Auction-based Advertisements: We derive MoVInG in the auction scenario described below to capture artificial competition.

- Second price auction (or Vickrey auction) [15], where each impression is auctioned off separately. From the revenue equivalence theorem, the expected revenue at equilibrium is equal to the expected revenue from a first price auction. In a second price auction, the winner is the bidder who bids the highest amount, but only pays the second highest bidder's bid amount. This incentivizes bidders to bid their true valuations [13].
- Bids drawn form uniform distribution $\mathcal{U}(0, r)$ (relaxed later). This assumption has been used in [12, 3].
- When targeting is based on a less precise data-set D', the conversion rate decreases proportionally to $\frac{S}{S'}$. Although in general advertisers may resort to very complex strategies to estimate the value of their ads (and to formulate their bids), it is fair to assume that a decrease in conversion rate will be translated in a corresponding decrease in bids: $b' = b * \frac{S}{S'}$. This does not affect the value and generality of MoVInG as it can be easily modified to accommodate the actual function that the advertisers use to set their bids.

In this auction scenario, the ad-networks' revenue is the expected *second maximum* from the bid distribution. Since the actual bids are drawn from a uniform distribution in the range $[0, r]$ (*i.e.*, $\mathcal{U}(0, r)$), the expected second maximum is a simple derivation known from [11]. For D this results in a cost per impression $c = \frac{n-1}{n+1} * r$, where n is the average number of advertisers competing for an impression, r is the maximum possible bid of an advertiser. Let the total number of advertisers in the system be A and total number of users in the system be U, the average competition per impression is $n = \frac{A*S}{U}$. Substituting in above equation, $c = \frac{A*S-U}{A*S+U} * r$. Similarly, with D': $c' = \frac{A*S'-U}{A*S'+U} * r'$, where $r' = r * \frac{S}{S'}$ is the new upper limit for the range of bids. Thus $c' = \frac{A*S'-U}{A*S'+U} * r * \frac{S}{S'}$. Hence the loss in revenue for an ad-network as given by MoVInG is:

$$L = c - c' = r * (\frac{A * S - U}{A * S + U} - \frac{A * S' - U}{A * S' + U} * \frac{S}{S'}) \quad (2)$$

Results: Here we compare MoVInG with fixed price ad placement (No Auction curve) and with simulations. Figure 3 and the following ones plot the *fraction* of revenue lost by ad-networks, *i.e.*, the loss expression in Eq. 1 (No Auction) or Eq. 2 (MoVInG) or loss obtained in simulations, divided by the original revenue with data-set D (obtained in that particular case). We set $A = 100, U = 1000, S = 100$, and vary S'. The simulation curves capture the contention among the A advertisers by emulating their bidding behavior as described in Figure 2.

```
Pick U users
Pick A advertisers and assign bids picked from the chosen distribution
(e.g., uniform distribution in Fig 3)
  for each advertiser:
      assign S number of interested users (for data-set D)
      assign S' number of interested users (for data-set D')
Create data structure B, B'
(for each user, all the interested advertisers for data-sets D, D')
Create payment sets P and P'
for each user in B:
      P[user] = second highest bid
for each user in B':
      P'[user] = second highest bid
L = average(P) − average(P')
```

Figure 2: Pseudo code for simulation.

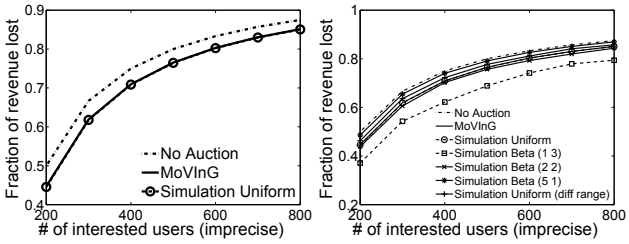


Figure 3: Uniform Distribution. **Figure 4: Varying Distributions.**

In Figure 3 the Simulation Uniform curve overlaps with the MoVInG curve, which demonstrates the correctness of the model. The estimated loss for ad-networks is $3\% - 12\%$ higher when increased contention is not captured. This is expected because increased contention increases the revenue for the auctioneer, thus reducing the loss. Parameter values in real world scenario may be higher than the ones used in our numerical evaluation; however, this does not cause any loss of generality because the results only depend on the relative values of the parameters. We also study the effect of varying A, S and U, but omit the results in the interest of space.

Figure 4 shows the loss in revenue (using simulations) when bids are picked from distributions other than uniform. The shape of the loss functions are similar, but just shifted depending on the bid distribution. Three Beta distributions $\mathcal{B}(\alpha, \beta)$ are considered with r = 1: (i) $\mathcal{B}(1,3)$, (ii) $\mathcal{B}(2,2)$, and (iii) $\mathcal{B}(5,1)$. The loss is higher (48.6% with $\mathcal{B}(5,1)$ vs 37% with $\mathcal{B}(1,3)$ at $S' = 200$) when distribution is skewed towards upper limit (*i.e.*, $\alpha > \beta$) as the initial revenue from such a bid distribution is very high. No Auction still has highest loss estimate: 50% at $S' = 200$. A case where advertisers draw bids from different ranges is also considered. The lower limit and upper limit of the range are picked, for each advertiser, uniformly between 0 and 1 and a uniform distribution within this range is used to draw bids. The result is similar to the one obtained with all advertisers picking bids according to a uniform distribution with the same range (45% with same range and 46% with different ranges at $S' = 200$).

Resemblance to Monotone Hazard Rate curves: The Equations 1 and 2 are forms of $1 - \frac{1}{x}$, thus have a decreasing slope and look like MHR curves which arise in a number of applications [4].

2.2 Generalized Expression

Here we derive an expression, for general bidding distributions.

Methodology: We derive the expression by fitting a large number of simulations run for different scenarios. From Eq. 2 under uniform bidding, and furthermore from the MHR shape of the simulation curves, we find that the loss L is a polynomial function in the following variables: S, S', A and U. Our goal is to have a model which is: (i) general and applicable to a variety of scenarios, (ii) has low error, and (iii) is simple and does not over-fit the data.

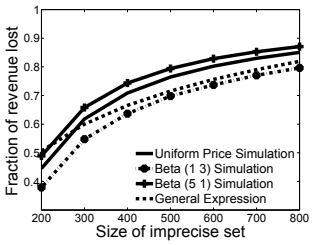


Figure 5: Accuracy of general expression.

We run simulations for varying A, S, S' and U giving us 189 data points. We further use different bid distributions: (i) Uniform $\mathcal{U}(0,1)$, (ii) Beta $\mathcal{B}(1,3)$, $\mathcal{B}(2,2)$, and $\mathcal{B}(5,1)$ and (iii) Uniform, bidders have different ranges, upper and lower limits picked uniformly between 0 and 1. This gives total of 945 data points to fit the model. We then write a general polynomial as

$$\begin{aligned} L \;=\; & x_1 U^{x_2} + x_3 A^{x_4} + x_5 S^{x_6} + x_7 S'^{x_8} + \\ & x_9 U^{x_{10}} A^{x_{11}} + ... + \\ & x_{27} U^{x_{28}} A^{x_{29}} S^{x_{30}} + ... + \\ & x_{43} U^{x_{44}} A^{x_{45}} S^{x_{46}} S'^{x_{47}} + x_{48} \end{aligned}$$

where x_i's are the unknowns and i is the index of each unknown. The unknowns include coefficients and exponents.

We solve a non-linear optimization problem to minimize the sum of squares of residuals. Once we obtain the coefficients and the exponents, we add additional terms to see if that helps accuracy (*e.g.*, multiple exponents per term). We find that this does not improve the accuracy. Thus, for simplicity, we use above equation.

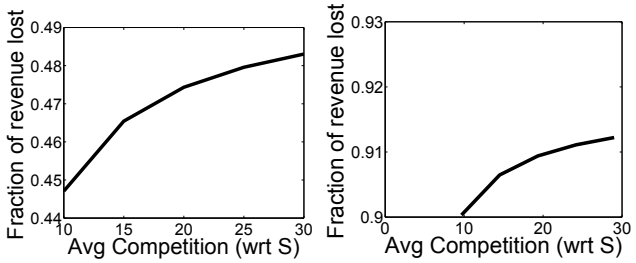
The expression thus obtained is still quite complex and may overfit for the simulation results. To further simplify it, we construct a linear system of equations and solve for x in $M * x = b$. Here M is a matrix of terms where rows correspond to different observations and columns correspond to different terms with exponents given by the above step, b is a vector of the real observed losses in simulations and x is a vector of unknowns specifying the *new coefficients* of the terms. We try to obtain a sparse solution for x through ℓ^1 norm minimization [1]. As shown in [9], the minimal ℓ^1 norm solution often coincides with the most sparse solution for many large linear systems. This avoids over-fitting the data and identifies most important factors. Thus we obtain final general model as:

$$\begin{aligned} L \;=\; & 0.40 * U^{0.17} + \mathbf{1.69} * A^{0.11} + 0.02 * S^{0.11} + \\ & 0.17 * A^{0.11} * S^{0.11} - \\ & 0.17 * \frac{U^{0.20} * A^{0.13}}{S'^{0.01}} - 0.07 * \frac{U^{0.20} * S^{0.14}}{S'^{0.01}} - \\ & \mathbf{1.60} * \frac{A^{0.10} * S^{0.10}}{S'^{0.07}} + \\ & 0.02 * U^{0.25} * A^{0.16} * S^{0.16} * S'^{0.04} \end{aligned}$$

Note that one of the significant terms (in bold face) contains $\frac{S}{S'}$ which is not surprising as the Eq. 2 contains the same term.

Accuracy of the general expression: In Figure 5 we plot the curve from the general expression along with the other simulation curves considered for the fitting. This is with our default parameters. Even though we look for a sparse solution, the accuracy of the model is very good (*i.e.*, within the range of $\mathcal{B}$ distributions).

We further quantify the error of the model, by evaluating scenarios other than the ones considered for fitting. We once again vary



(a) Gender (b) Age

Figure 6: Loss in revenue in real data-set with simple queries.

A, S, S' and U to give 64 data points (different from ones used for fitting). Further we run this for several bid distributions: (i) $\mathcal{U}(0,1)$, (ii) $\mathcal{B}(0.5,0.5)$ and (iii) $\mathcal{B}(2,5)$. We observe that the general expression, captures the loss in revenue accurately with 6.8% mean relative error ($= \frac{1}{k} * \sum_{i=1}^{k} \frac{abs(actual\_loss_i - estimated\_loss_i)}{actual\_loss_i}$) with $k = 64 * 3 = 192$ trials.

3. TRACE DRIVEN EVALUATION

Here, we use MoVInG on a real data-set. We assume the full data-set from a large CSP to be the most precise data-set D. We construct synthetic D's assuming some attributes to be absent. We first consider simple queries, *i.e.*, queries on *single* attributes.

Gender: Given 18024 users with gender information, for varying number of advertisers, we assume 50% of advertisers are interested in advertising to males and the rest 50% to females. If gender information is available (D), average size of users satisfying the query $S = \frac{18024}{2} = 9012$ and if gender is unavailable (D'), $S' = 18,024$.

In Figure 6 (a), we plot the loss in revenue with varying levels of average competition. We define average competition as $\frac{A*S}{U}$, *i.e.*, the average number of advertisers, interested in each impression auctioned. The fact that loss of revenue with average competition of 10 is already over 0.44 means that the loss of gender information is already cutting the revenue almost by half (it is not exactly at 0.5, because the effect of increased artificial competition compensates for some loss). Moreover, fractional loss increases with increase in real competition, because at higher competition levels, artificial competition doesn't help as much.

Age: We vary average competition as above and assume that advertisers are interested in users of any age with equal probability. From the data-set we have age information of $14,348$ users. We divide them into 6 age groups 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. D is the data-set with exact age information available: S = 196.5 and D' is the data-set without age information available but age group information available, $S' = 2,391.33$. In Figure 6 (b), we plot the loss in revenue comparing S and S'. We see that loss is over 90% because the difference between S and S' is large.

Composite Queries: Here, we consider an example composite query, *i.e.*, query on *multiple* attributes, *e.g.*, a jeweler wants to advertise to all users that are female and in age group $25-34$, *i.e.*, query $Q = (25 \leq \mathrm{v}(a_{age}) < 35 \wedge \mathrm{v}(a_{gender}) = Female)$.

Suppose we have advertisers, each interested in one of the 12 queries in Table 2. Also suppose universe is the people who have age and gender information in our data-set: $U = 13,598$ with males = $8,612$ and females = $4,986$. Thus average set size S'_{no_age} if no age group information is $6,799$. From the table, if both age group and gender are available, average set size $S = 1,133$ and average set size S'_{no_gender} if only age group is available is $2,266$. Fraction of revenue lost, with S'_{no_gender} = 0.45 and with S'_{no_age} = 0.80 (with average competition with respect to precise set $n = 10$).

Queries	Gender	Age Group	Count	Total Count
Q_1	Male	1	2716	4450
Q_2	Female	1	1734	
Q_3	Male	2	2423	4317
Q_4	Female	2	1894	
Q_5	Male	3	2559	3517
Q_6	Female	3	958	
Q_7	Male	4	769	1106
Q_8	Female	4	337	
Q_9	Male	5	118	169
Q_{10}	Female	5	51	
Q_{11}	Male	6	27	39
Q_{12}	Female	6	12	

Table 2: Composite query on gender and age.

As expected, the loss is higher if age group is not available, than when gender is not available as the S'_{no_age} is $6\times$ larger than S and S'_{no_gender} is only $2\times$ larger than S.

4. FUTURE DIRECTIONS

Two directions will be followed in future work: first different from the assumption in this work, the ad-network may use a strategy wherein they do not include a user in the returned set if the information required by the query is unavailable. Thus user-set returned maybe smaller in case of D' than in case of D. It will be interesting to study the mix of these two strategies. Second, we have ignored the strategic aspects of bidding. However, what will be the bidding strategy of an advertiser interested in D when she is only allowed to target D'? Recent research on game theoretic and equilibrium analysis in presence of auctions with information [10, 2] could be used as a starting point to extend our work.

5. REFERENCES

[1] A method for large-scale l1-regularized least squares. `http://www.stanford.edu/~boyd/l1_ls/`.

[2] I. Abraham, S. Athey, M. Babaioff, and M. Grubb. Peaches, Lemons, and Cookies: Designing Auction Markets with Dispersed Information. 2011.

[3] R. Bapna, P. Goes, A. Gupta, and G. Karuga. Optimal design of the online auction channel: Analytical, empirical and computational insights. *DECISION SCIENCES*, 33:557–577, 2002.

[4] R. E. Barlow, A. W. Marshall, and F. Proschan. Properties of Probability Distributions with Monotone Hazard Rate. *The Annals of Mathematical Statistics*, 34:375–389, 1963.

[5] Behavioral targeting conversion rate. `http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf`.

[6] Behavioral targeting spending. `http://www.emarketer.com/Article.aspx?R=1006384`.

[7] Behavioral targeting usage. `http://www.internetretailer.com/2009/05/04/24-of-online-marketers-use-behavioral-\\targeting-up-50-from-a`.

[8] Data trade. `http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf`.

[9] D. L. Donoho. For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution. Technical report, Comm. Pure Appl. Math, 2004.

[10] Y. Emek, M. Feldman, I. Gamzu, R. P. Leme, and M. Tennenholtz. Signaling schemes for revenue maximization. In *ACM Conference on Electronic Commerce*, pages 514–531, 2012.

[11] Expected second maximum. `http://www.econ.ucsb.edu/~tedb/Courses/Ec100C/Readings/MatthewsAuctionPrimer.pdf`.

[12] A. Ghosh, R. P. McAfee, K. Papineni, and S. Vassilvitskii. Bidding for representative allocations for display advertising. *CoRR*, abs/0910.0880, 2009.

[13] V. Krishna. *Auction theory*. Academic Press, 2002.

[14] S. Muthukrishnan. How to auction data. `https://sites.google.com/site/adauctions2011/program-and-schedule`.

[15] W. Vickrey. Counterspeculation, Auctions and Competitive Sealed Tenders. *Journal of Finance*, pages 8–37, 1961.