# Scalable Online Ad Serving: Experimental Comparison of Simple Techniques

Gang Wu and Brendan Kitts
Microsoft adCenter
One Microsoft Way
Redmond, WA 98052
+14257077219
{simonwu,bkitts}@microsoft.com

## ABSTRACT

Online Ad Servers attempt to find the best ad to serve for a given triggering user event. The performance of the ad may be measured in several ways. We suggest a formulation in which the ad network is trying to maximize revenue subject to relevance constraints. We describe several simple algorithms for ad selection and review their complexity. We tested these algorithms using the Microsoft ad network from October 1 2006 to February 8 2007. Over 3 billion impressions, 9 million combinations of triggers with ads, and a number of algorithms were tested over this period. We discover curious differences between ad-servers aimed at revenue versus clickthrough rate.

## Categories and Subject Descriptors

I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems

## General Terms

Algorithms, Experimentation.

## Keywords

Ad serving, Online advertising.

## 1. INTRODUCTION

Online Advertising networks such as Doubleclick [2] and MSN [1] serve ads to users visiting web pages. Ad inventory typically comprises millions of different ad creatives, each of which have their own unique constraints and agreements for payment.

Constrained ad delivery has been discussed in detail by [5]. In this paper we describe an ad serving application with fewer constraints, but focus on the core problem of serving high revenue ads, and estimating ad performance statistically.

We also diverge from the literature in modifying the objective function. Typically ad serving is conceptualized as a problem of serving the ad which will generate the highest revenue [4][6]. We argue that this kind of approach has led to a proliferation of irrelevant and financially-orientated ads online – for instance, ads about "mortgage" and "loans", and "refinancing". We show examples of this in the final section.

We instead formulate the objective function as one of maximizing revenue subject to well-defined relevance constraints. This approach allows the ad server to maintain minimal standards of user experience in serving its ads.

We tested these algorithms live from October 1 2006 to February 8 2007. Over 3 billion impressions, 744,642 unique ad-trigger pairs, and 6 algorithms were tested over this period. We discover that approximate methods of estimating ad clickthrough rate are effective in production, smoothed predictions raise optimization performance although modeling performance is lower, and that explicit optimization methods are also effective but more expensive.

## 2. ONLINE ADVERTISING

A trigger $t$ is any user-initiated event such as a webpage impression, typed search query, length-of-time-on-page, a user with a particular profile visiting a page, or any other behavior that may be valuable to marketers. Online advertising is powerful because the triggering events are used to provide context for the display of the advertisement. Conversion rates as high as 80% are possible for well-targeted advertisements, particularly those in query search marketing [9].

In response to a trigger, online advertisements are served to the user. If the user takes some action after their exposure to the ad such as clicking on the ad, or converting with the advertiser, then a payment event is generated and the publisher is paid by the advertiser. Revenue events commonly *Pay per click*, *Pay per impression* and *Pay per acquisition*.

The trigger, advertisement, revenue event relationship can be generalized as a variable $r_{k,t}$ where $t$ is the triggering event and $k$ is the advertisement that is displayed to the user and $r$ is the amount the advertiser agrees to pay if a subsequent behavior is observed from the user.

***Example.*** Say that Joe is advertising shoes on publisher everythingathletic.com. The triggering event is a user viewing a page everythingathletic.com/athleticfootwear.html. The advertisement Joe might want to show is about Nike Airmax shoes. The revenue event is a click of the user on Joe's advertisement – if this occurs, Joe will pay a pre-arranged amount $0.10.

## 3. THE AD SERVING PROBLEM

The ad server has a large amount of flexibility in which ads it serves in response to a triggering event. The ad server has to respect the advertiser's pre-arranged requirements $r_{k,t}$, which indicates that $k$ may be displayed in response to $t$. However, the ad server will want to rank the different ads in a variety of ways so as to achieve ad-network revenue and relevance goals.

An ad-server function $I_{k,t}$ indicates whether ad $k$ is displayed in response to trigger $t$. It is important that an advertisement be served back to the user that is optimal in terms of maximizing revenue for the advertiser, for clickthrough rate (CTR), for user relevance, or some other well defined metric.

***Definition*:** Revenue maximization subject to relevance

Given a set of revenue agreements $r_{k,t}$, find an ad selection for each trigger, $I_{k,t}$ such that when $I_{k,t}$=1, ad $k$ is displayed in response to trigger $t$, and no other ads are displayed. Such ad selections $I_{k,t}$ should be chosen so as to maximize ad-server revenue subject to relevance constraints.

$$I_{t,k} \in [0,1] : \max \sum I_{k,t} r_{k,t} c_{k,t} \qquad (1)$$

$$\text{subject to} \sum_{k} I_{k,t} \leq P \qquad (2)$$

$$\text{and} \sum_{k,t} I_{k,t} \leq N \qquad (3)$$

$$\text{and} \sum I_{k,t} c_{k,t} \geq CTR_{\min} \qquad (4)$$

(1) is a revenue function. In this function we see three variables: $c_{k,t}$ is the probability of the payment event and $r_{k,t}$ is the agreed upon revenue to be paid from the advertiser if a payment event is detected, and $I_{k,t}$ is the ad-serving function which determines which of the advertisements are shown in response to a trigger $t$.

(2) is an ad-delivery constraint. We were able to deliver $P$ or fewer ads per trigger, and we were able to refuse to deliver ads for certain triggers if this would have a poor result on revenue or relevance. In other applications all $P$ ads may need to be delivered so as to fit into an iFrame, and so the <= would change to an = equality for these applications. We have noted that in other ad-serving contexts, constraints are considerably richer and may include constraints on size of ad, time of day, and so on.

(3) is a maximum number of trigger,ad pairs that will be stored by the ad-serving system. This is a practical constraint, and can be set to infinity if the ad-server is able to service an unlimited number of trigger,ad pairs.

(4) is a relevance constraint indicating that the ads delivered are meeting a minimum level of user experience. Many metrics may be used to measure relevance of advertising including revisit probability and clickthrough rate, the latter of which is reflected in the formula above.

$c_{k,t}$ is the only variable which is unknown. A method for predicting trigger-ad-paymentevent probabilities from historical data is therefore needed to so that the objective function can be maximized. The next section will describe robust and computationally efficient methods for estimating the trigger-ad probabilities.

## 4. CTR PREDICTION

In order to maximize revenue the unknown is the probability of the payment event $c_{k,t}$ and this must be estimated statistically. The problem with this task is the cross-product of $T \times K$ is extremely large and sparse. In our empirical data, out of 5 million trigger-ad possibilities, less than 2%, had any data about click in response to a trigger,ad pair – this included single impressions with zero clicks.

For pairs without any impressions – or with low and potentially statistically unreliable numbers of impressions, it is necessary to estimate probability of click using some additional information.

We investigated a variety of approaches for predicting payment probability. These are described below.

### 4.1 Expected CTR of Advertisement

This algorithm (labeled "globalctr" in the figures) works by setting the probability of click given a trigger,ad equal to the probability of click given the ad summed over all triggers. For instance, if customers tend to click after seeing an advertisement for Britney Spears music CD, the ad server would tend to assume that Britney Spears music CD advertisements will generate a click, even if the triggering page is something unrelated, eg. SCUBA diving.

$$\Pr(k) = \sum_{t} \frac{C_{t,k}}{IMP_{t,k}} \qquad (5)$$

This algorithm does not take into account the particular context of the user. Although this may seem poor, it has certain advantages.

(a) Low Computational complexity: Algorithm space scales with the number of ads O($K$) rather than with the product of ads with triggers. The CTR statistics themselves can be computed during a single pass, with accumulators for clicks, impressions, and so this scales very well on large amounts of data.

(b) Statistics for ads are fairly dense, and so estimation is relatively easy.

### 4.2 Expected Conditional CTR

This method (labeled "history" in the figures) involves using the empirical counts of clicks and impressions to estimate conditional CTR for each pair. A count of clicks given ($t$,$k$) pair, and impressions of ($t$,$k$) pairs is taken, and then the two are divided to generate conditional clickthrough rate. Accumulators can be used to record clicks and impressions for each observed ($t$,$k$) pair. However, space complexity is now O($T$*$K$), and this may be impractical in some ad serving domains.

A further disadvantage is that the sparsity leads to over-generalization when very low samples are available, eg. if 1 impression was observed and 1 click, the CTR is estimated at 100%, or 1 impression and 0 clicks then a CTR of 0% is estimated. The vast majority of trigger-ad pairs are likely to have never received an impression, nor generated a click. In our data, just 2% of all trigger-ad pairs had impressions data.

$$c_{t,k}[2] = \Pr(k \mid t) = \frac{C_{t,k}}{IMP_{t,k}} \qquad (6)$$

## 4.3 Bayesian prior combination with observed conditional CTR

This method (labeled "smoothed" in the figures) utilizes the observed conditional clickthrough rates, but weights in an *a priori* estimate of the clickthrough rate of the group of ads. This approach has been described elsewhere in the literature [7][3]

$$\Pr(k \mid t) \cdot f(IMP_{k,t}) + \Pr(a) \cdot \left[1 - f(IMP_{k,t})\right],$$

$$\text{where } f(IMP_{k,t}) = 1 - e^{-\alpha \cdot IMP_{k,t}}$$

$$\text{and } \alpha = 0.01 \qquad (7)$$

In our application trigger-ads were grouped into algorithm buckets $a$, and these had a certain clickthrough rate $\Pr(a)$. As a result, we could use the observed conditional clickthrough rate, and weigh in the algorithm clickthrough rate.

The weighting of the observed clickthrough rate increased with the number of impressions, and at 100 impressions more weight was put onto the observed pair's clickthrough rate than on the algorithm's expectation for clickthrough rate.

## 4.4 CTR Prediction based on features

We tested both linear regression ("linear" in the figures) and decision tree ("dtree" in the figures) for predicting clickthrough rate. The decision tree was similar to C4.5 and has been described in [8].

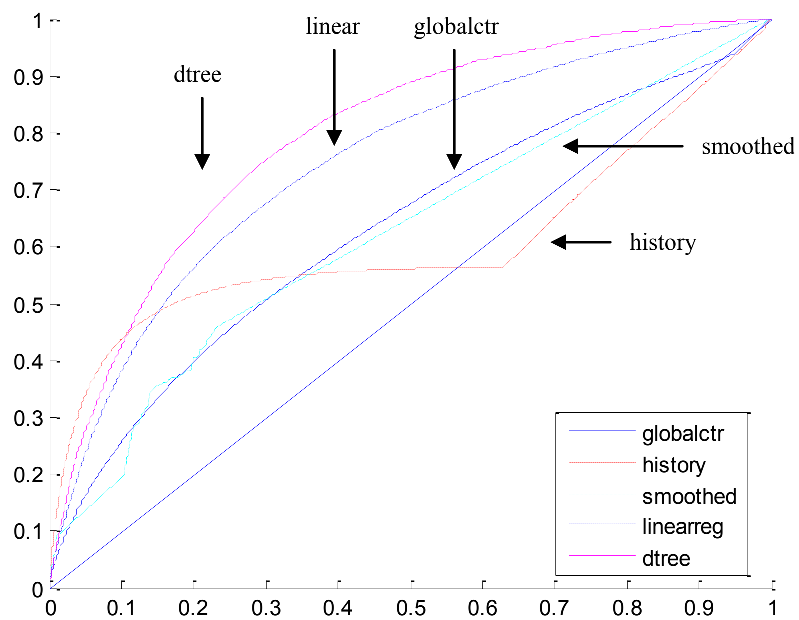


**Figure 1. Response Operator Curve to predict CTR > 0.09, five algorithms shown**

**Table 1. Clickthrough rate prediction accuracy for five algorithms**

| | Globalctr Cost | history Cost | Smoothed Cost | linear Cost | Dtree Cost | Globalctr CTR | history CTR | smoothed CTR | linear CTR | dtree CTR |
|---|---|---|---|---|---|---|---|---|---|---|
| 25th | -$0.046 | -$0.031 | $0.001 | -$0.100 | -$0.032 | -2.44% | -0.40% | 0.26% | -2.19% | -1.92% |
| 90th | $0.021 | $0.022 | $0.086 | $0.070 | $0.028 | 5.15% | 6.25% | 7.14% | 4.10% | 3.73% |
| 10th | -$0.239 | -$0.188 | -$0.050 | -$0.157 | -$0.117 | -5.36% | -2.53% | -16.16% | -3.28% | -3.19% |
| 75th | $0.003 | $0.005 | $0.021 | $0.020 | -$0.001 | 1.65% | 2.63% | 3.53% | 0.91% | 0.78% |
| Rankcorr | | | | | | 0.236 | 0.208 | 0.252 | 0.427 | 0.541 |
| Corr | | | | | | 0.254 | 0.417 | 0.237 | 0.388 | 0.452 |
| AUC | | | | | | 0.629 | 0.605 | 0.621 | 0.748 | 0.794 |

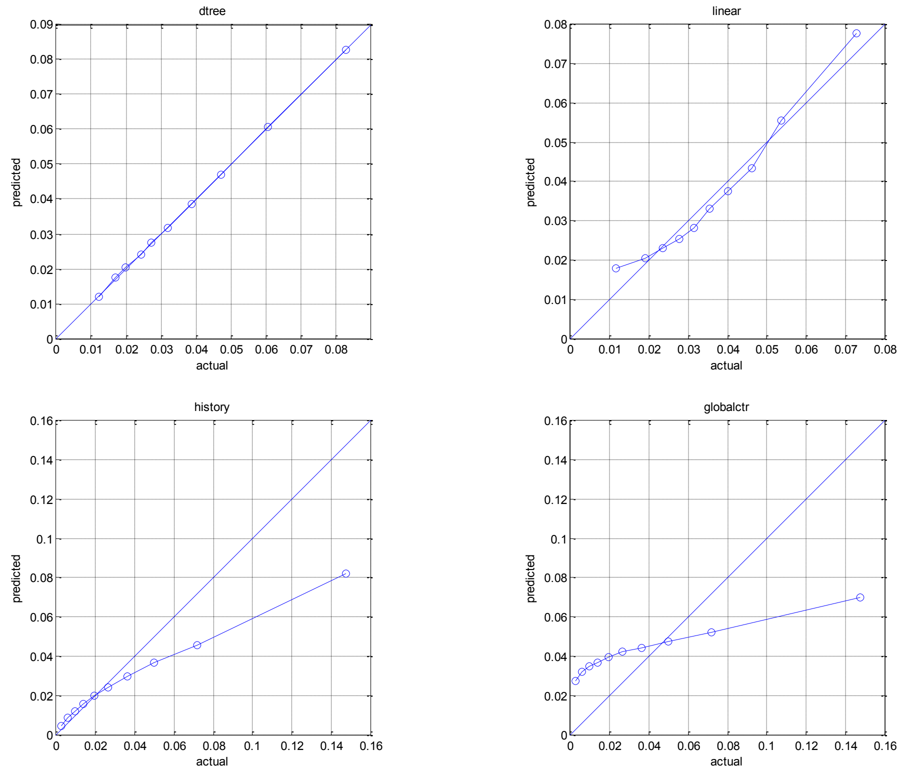


**Figure 2. Clickthrough rate prediction for 10 actual deciles (x-axis) against aggregated predicted clickthrough rate (y-axis). Perfect behavior would have the predictions matching the actuals exactly, and so falling along the same diagonal line.**

## 4.5 Clickthrough rate prediction results

We used 3 months of data on trigger-ad displays and clickthroughs from October to December 2006, to predict the following month of trigger-ad clickthrough results in January 2007. All algorithms were scored on the same trigger-ad pairs.

Table 1 shows Spearman's correlation metric for each of the methods. This correlation statistic ranges from about 0.2 for the globalCTR, historicalCTR, smoothed methods, to a maximum of about 0.5 for the decision tree approach.

Response Operator Curves were also created by discretizing the target clickthrough prediction into "high" clickthrough rate cases which had clickthrough rates greater than 9% (Figure 1). ROC curve shows that the historical-conditional-ctr method generates very good performance initially, however, ultimately the method falls victim to a missing value problem. This is because many trigger-ad pairs that are not encountered in the training period, but generate clickthroughs in the future period. Thus, the approach shows good performance initially as pairs which have been observed are selected, but then shows random performance as cases which were 0 in the historical period essentially predict 0 in the future period.

Figure 2 shows a comparison of the algorithms across different deciles of actual clickthrough rate values. The decision tree approach out-performs the other methods.

# 5. AD SERVER OPTIMIZATION

We next turn to the task of solving the objective function so as to find ad selections $I_{k,t}$.

## 5.1 Greedy Ad Serving Algorithm

In order to cope with the scale, we adopted the greedy algorithm of selecting the highest value to impressions ratio ads that satisfied the serving constraints. This algorithm is well known for knapsack problems. Because the optimization objective was clickthrough rate, this meant in practice selecting the highest (clickthrough rate, number of searches) ads in order of first variable, second variable descending.

## 5.2 Closeness to Optimality

The Maximum difference between optimal solution and generated solution under greedy CTR allocation is O(max $IMP * CTR$) where max $IMP*CTR$ is the maximum number of clicks per day from any trigger-advertisement. The sketch of this proof is that selecting the item with highest clicks to impressions ratio necessarily is the best item to add for a given level of cost (impressions), will hold recursively until we reach a final item that is suboptimal. The final item will have the property that when we add it, there is another item which could have been added which would have perfectly filled our knapsack up to the desired number of impressions. Thus we need only find the item with largest benefit as the upper bound on the distance from optimality.

Figures 5 shows distance to optimal solution for different scales of solution. As more trigger-ads are added,

Empirically the maximal benefit item was 138,370 clicks. As a result, if we add the top 1,000 ads, the worst-case distance to optimality is 75 times the number of clicks generated under the greedy solution. At 50,000 pairs, number of clicks being lost due to greedy drops to just 39% of the quantity obtained under the greedy solution.

Figure 3 and 4 show cumulative clicks versus clickthrough rate, versus distance to optimal solution as more and more trigger-ads are loaded into the system.

### 5.3 Revenue versus Clickthrough rate

In order to examine the difference between ad-serving for revenue maximization versus ad-serving to achieve the highest clickthrough-rate\, we chose to test both optimization extremes, and report on the result here. In order to achieve clicks optimization, we set $r_{k,t}=1 \ \forall \ k,t$; otherwise we used the objective function in section 3 unchanged.

## 6. EXPERIMENTAL RESULTS

We tested these algorithms live from October 1 2006 to February 8 2007. Over 3 billion impressions, 744,642 unique ad-trigger pairs, and 6 algorithms were tested over this period. We deployed the following ad-serving algorithms:

We noted in Section 4 that the best clickthrough rate predictors (aside from decision tree and linear regression methods) were roughly historical, followed by global, followed by smooth generate The comparison of ad-serving performance using each of these predictors to inform the clickthrough rate is shown in Figure 5 (left). Historical conditional clickthrough rate generated the highest global clickthrough rate, followed by smoothed. This is consistent with the higher accuracy of the historical method. We now turn to the relationship between clickthrough rate and revenue.

Figure 5 (right) shows the relationship between the clickthrough rate of an ad and revenue produced by the ad. Higher clickthrough ads produce more revenue up until the 80th percentile of the clickthrough rates. After that, revenue produced by higher-clickthrough ads actually declines. This means that there may be "hyper-relevant" ads which generate less revenue. The inverted "U" function has not to our knowledge been noted before.

Table 2 shows the top trigger-ads selected for clickthrough rate maximization versus revenue maximization. Revenue maximizer selects a disproportionate number of financially-related terms including "refinance", "loan", "mortgage", and so on. This bias is not apparent in the case of clickthrough rate maximization. We conjecture that this is why banner advertisements are often contextually unrelated to the page they are on, and are dominated by mortgate, loan, and refinancing creatives.

## 7. CONCLUSION

We have presented a general framework for ad serving and have tested clickthrough-rate prediction algorithms experimentally in the Microsoft ad network. These clickthrough rate prediction algorithms were tested experimentally within an ad-serving context, and we were able to show that higher accuracy algorithms do generate more clicks when exposed to human traffic. We also experimentally tested extreme optimization solutions – one maximizing revenue, and the other clickthrough rate – and have noted interesting differences between the results produced. Our results suggest that the bias towards revenue-generating advertisements needs to be tempered with well-defined relevance constraints.

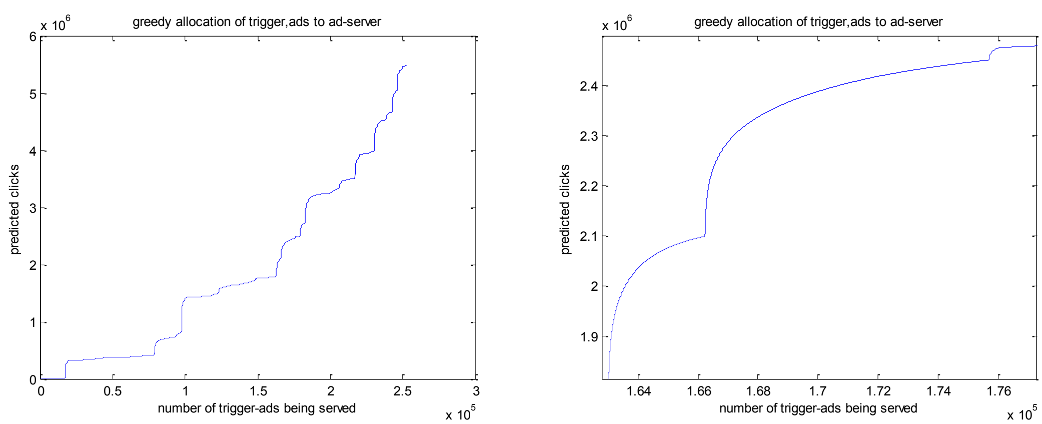


**Figure 3. (left) Greedy optimization to achieve highest clickthrough rate. Trigger-ads are added in order of clickthrough rate, clicks descending. As a result, the highest clicks items are added first, and then lower clicks. Because a decision tree was used with different discrete click levels, this results in a banding pattern as seen above, although other models could have also produced this banding if equal clickthrough rates were generated. Despite this, the Clickthrough rate is monotonically decreasing, which is needed for the greedy property to work. (right) close-up showing a clickthrough rate band**

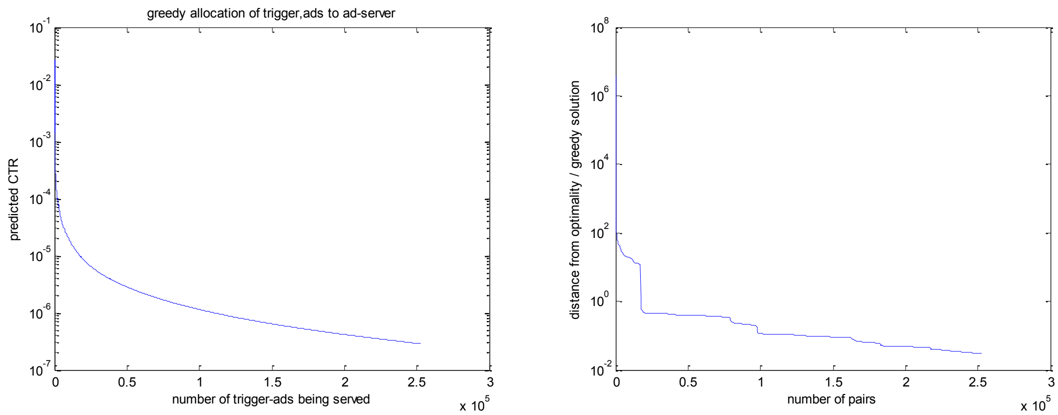


**Figure 4. (left) Clickthrough rate versus number of trigger-ads being served. Clickthrough rate decreases monotonically as more trigger,ad pairs are added. (right) Scalar distance between greedy solution and optimal solution, where distance is equal to worst case optimal solution / greedy solutin.**

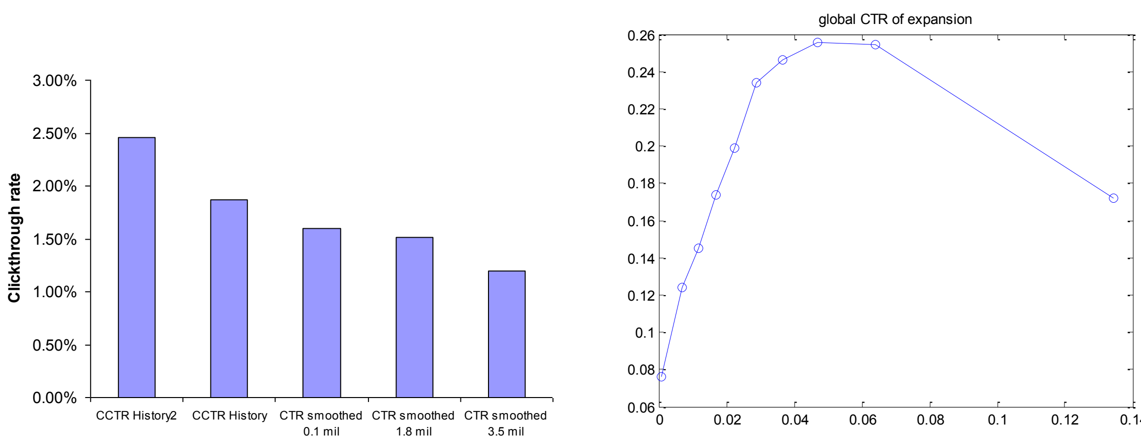


**Figure 5. (left) Ad-serving results for clickthrough rate maximization, using one of three algorithms. Algorithms that showed better ability to predict CTR tended to do better in generating clicks in ad-serving. (right) global CTR of advertisement versus revenue generated per display. Higher clickthrough rate ads generally make higher revenue, however, ads with very high clickthrough rates generate only half the revenue available**

**Table 2. Top revenue versus Top CTR trigger,ads. Revenue optimization results in a large number of financially related terms being selected (shaded). Clickthrough rate optimization does not show this kind of bias. We conjecture that this is why banner advertisements are often contextually unrelated to the page they are on, and are dominated by mortgate, loan, and refinancing creatives.**

| **Rank** | **Objective=Revenue** | | **Objective=Clickthrough rate** | |
|---|---|---|---|---|
| | ***Trigger*** | ***Advertisement*** | ***Trigger*** | ***Advertisement*** |
| 1 | renu | renu eye infection | Flyingflowers | flying flowers |
| 2 | whole life insurance no medical exam | life insurance no medical exam | Matlin | marlin |
| 3 | guaranteed faxless payday loan | guaranteed faxless payday loans | hotter com | hotter shoes |
| 4 | mobile internet service | mobile satellite internet | www american airlines | american airline |
| 5 | personal loans bad credit | bad credit loans | jackson perkins roses | jackson perkin rose |
| 6 | mortgages bad credit | mortgage bad credit | Barnesandnobel | barnes nobel com |
| 7 | online roulette | roulette online | west elm | www westelm com |

| | | | | |
|---|---|---|---|---|
| 8 | charity cars | charity car | maplins electronics | maplin electronics |
| 9 | wheelchair lift vans | handicap lift | free online ganes | free online games |
| 10 | customer relationship manager | customer relationship management | toyota solaro | toyota solara |
| 11 | instant life insurance quotes | instant life insurance quote | Matlins | marlins |
| 12 | 2nd mortgages | 2nd mortgage rate | alligient airlines | allegiant airlines |
| 13 | compare mortgage rates | best interest rates | oriental training | oriental trading |
| 14 | broadband phone service | dsl broadband | Earlylearningcentre | early learning centre |
| 15 | auto insurnace | auto insurance | premier travel inns | premier travel lodge |
| 16 | www insurance companies | insurance company | ll bean kids | llbean com |
| 17 | chevy avalanche accesories | chevy avalanche accessories | active hotels | active hotel |
| 18 | nextstudent com | college loans | stockmans bank Arizona | stockman bank arizona |
| 19 | hp printer tech support | hp printer | crates barrel | crate barrell |
| 20 | credit card consolodation | credit card consolidation | 101cds | 101cd |
| 21 | buy notes | buying note | loehmans plaza | loehman plaza |
| 22 | equity line credits | equity line credit | www famousfootwear com | famous footwear |
| 23 | nn125 home equity loans | home equity | http www ebay com | ebay home |
| 24 | paydayadvances | payday advance | jcpenney catalog | j c penney catalog |
| 25 | family planing | family planning | reverse look | reverse look up |
| 26 | best buy let mortgages | buy let mortgage advice | greatlakes com | great lakes |
| 27 | bad credit rate | bad credit rating | | |
| 28 | auto insurance com | car insurance | | |
| 29 | stocks investments | stock investing | | |
| 30 | stock trading companies | stock trading company | | |

## 8. ACKNOWLEDGMENTS

Thanks to Microsoft for making this research possible

## 9. REFERENCES


[1] What is Microsoft adCenter?(2007), Microsoft Corporate website, https://help.live.com/help.aspx?project=adcenter_keywords_LIVE_ss&market=en-us&querytype=keyword&query=77yek&tmt=&domain=adcenter.microsoft.com&format=b1

[2] Dart Motif: Ad Serving Features (2007), Doubleclick corporate site, http://www.dartmotif.com/formatsfeatures/formatsfeatures_adserving.asp

[3] Buhlmann, H. (1967), Experience rating and credibility, ASTIN Bulletin, Vol. 4, pp. 199-207.

[4] Ad serving. (2007, February 22). In Wikipedia, The Free Encyclopedia. Retrieved 19:06, February 25, 2007, from http://en.wikipedia.org/w/index.php?title=Ad_serving&oldid=110102191

[5] Amiri, A. and Menon, S. (2003), Efficient scheduling of Internet banner advertisements , ACM Transactions on Internet Technology, Vol. 3, Iss. 4, pp. 334-346. ACM Press.

[6] Dwight, Allen, Merriman, et al. (1999), Method of delivery, targeting, and measuring advertising over networks, USPTO Patent Number 5,948,061

[7] Hardy, M. (2007), Topics in Actuarial Analysis: Bayes, Buhlmann and Beyond, Financial Engineering News, Iss.. 45, http://www.fenews.com/fen45/topics_act_analysis/topics-in-act-analysis.htm

[8] Kitts, B. (1997), Regression Trees, Technical Report, http://www.appliedaisystems.com/papers/RegressionTrees.doc

[9] Kitts, B. Laxminarayan, P., LeBlanc, B. and Meech, R. (2005). A formal analysis of search auctions including predictions on click fraud and bidding tactics. ACM Conference on E-Commerce – Workshop on Sponsored Search, Vancouver, UK. June 2005. Available October 15, 2005, at http://research.yahoo.com/workshops/ssa2005/sched.html.