DOI: 10.20472/IAC.2018.039.042

KAPIL SHARMA

Delhi Technological University, India

TECHNOLOGICAL ASPECT IN REAL TIME BIDDING: A PROBABILISTIC APPROACH

Abstract:

This paper aim is to study the process of real time bidding or real time auctions for online digital advertising. Real time bidding drives the focus of bidding strategy from the user's profile by calculating a bid for each impression in real time. Real Time Bidding uses computers and multiple software's which implements multiple algorithms to display ads per impression via real time auction. It has been seen that by taking different parameters (e.g. conversion rates for a targeted audience), those account for varied prices at different market segments or pricing schemes. The data mining model implemented is the Statistical Arbitrage Mining (SAM). The campaigns use the CPA (cost per action) method on the meta-bidder to accomplish CPM (cost per mille-impressions) ad inventories paradigm thereby reducing the advertiser's risk. In SAM, trying to seek the optimal bidding price to maximize the expected arbitrage net profit is the net goal. A modern portfolio base is implemented to manage the risk. The Expectation - Maximization (EM) fashion is used to estimate the profit of each campaign and thereby maximize it. By using this, the meta-bidder successfully catches the statistical arbitrage opportunities in RTB. Also using the concepts of finance, the calculation of risk is done for each campaign.

Keywords:

Expectation - Maximization; Bidding; Statistical Arbitrage Mining

1.Introduction

Real Time Bidding marks the new era of digital advertising. It is similar to financial markets but follows an instantaneous programmatic approach. In this approach we bought advertising inventory and that is sold on the basis of per-impression basis. It is autonomous and completes a full transaction in milliseconds. The main advantage of Real-time bidding is that it lets advertisers to manage and optimize ads collected from multiple ad-networks. Users are granted multitude of various networks, which allows them to launch and create advertising campaigns. The other advantages are that users can prioritize networks. They are also allocated percentages of unsold inventory. This unsold inventory is known as backfill. RTB is also known as programmatic bidding. The advantage of programmatic bidding is to allow companies which represents sellers and buyers to bid. This bid is on the price to show an ad to a user. This bid happens every time a banner ad gets loaded. When a page or a site gets loaded thousands of bidders bids to serve an ad to that user. That bid is based on individual algorithms for each company. The buying and selling of the advertising inventory takes place on a technology platform known as ad-exchange. This is technology driven as opposed to the earlier idea of negotiating on media inventory. The bid is submitted within a very small amount of time, usually less than 100ms. Real time bidding has led to a significant Return On Investment(ROI) for the advertisers and focuses on behavioral targeting of user data.

The problem statement of this paper is how to determine the bid price while achieving advertiser target such that the net arbitrage profit is maximised while the risk involved is minimized (Both in a single campaign and across multiple campaigns).

A primary motivation for following RTB today is that is autonomous and takes place in milli-seconds. Online Advertising takes place on a large scale today and is a novel topic and thus it made us interested in learning about the various approaches and propose new ideas that can further optimize the process of real time bidding for online advertising. However, the biggest challenge involved is how to estimate and divide the budget across multiple campaigns by minimizing the losses and maximising the profits and also avoiding to exhaust the whole budget.

This rest of the paper has been organized as follows. Section 2 gives the literature review of this project Section 3 explains the methodology used in this project. Section 4 provides the results obtained, Section 5 includes conclusions and Section 6 introduces proposed future work. In the end references are also provided.

2.Realted Literature

In [1] , the whole process of real time bidding is explained which is different to the conventional asserting a predefined advertisement for a specified website . Real time bidding takes into consideration the user's profile and then an auction for the ads are done. This paper explains about the main components of the whole real time bidding process and how the elements are interlinked with each other. For an advertisement to be published on a website, the publisher side sends a request to the RTB exchange which further sends this to the Demand Side Platform (DSP). An auction takes place and whichever advertisement has the highest bid price wins it.

In [2] advertisers buy spots in ad exchanges. Before the reaching of the ad to the audience, it passes through many exchanges and data analysts. To avoid this overhead, the use of advertisers or campaigners are seen. Using their own algorithms to calculate bid prices, they ferret out underpriced ads and resell them.

In [3] , the features of an online exchange are mentioned . It explains how normal advertisements have different goals than online ads. The online ad process has an effect on both sides. It has to prove beneficial for the customer who is viewing it. If it is of relevance and the user clicks on the ad, profit is then only generated by the advertiser side. This is in correlation with the whole basic concept of online advertisements as per the need and choice of the user. Focusses on the goals of an exchange design such as efficiency, expression and simplicity.

In [4], the detailed statistical analysis of this dataset is presented. Real time bidding is an emerging technology which lacked publicly available datasets that could be used to conduct the experiments and create solutions for the real time bidding research problems. In the year 2013 iPinYou organized a real time bidding algorithm competition globally for the duration of three seasons. There were two stages in each season: one was the online stage and the other was the offline stage. A dataset for modelling training and testing later was kept in the offline stage. The dataset consists of number of clicks, number of conversions, final conversions and logs of Ad bidding etc. After the completion of the competition iPinYou released the dataset for public use for all the three seasons. [5] gives a detailed explanation of these datasets that could be used for conducting research in bidding strategies, CTR estimation etc. We have also utilized this dataset in our thesis.

In [6] , the impact of publisher determines the conversion rate or at least the probability of an advertisement to be converted. This helps in further bid optimization so that the appropriate price is placed for the advertisements.

While placing an advertisement on any impression it is imperative to focus on its ability to convert a user into a customer. Due to large amount of data available and uncertainty present in the user data, finding probability for conversion is a difficult task. [7] establishes an approach based on data hierarchical model, where the levels are user, publishers and, advertiser whose previous performances are being analyzed, to estimate the conversion rate. Different levels possess different binomial distribution and the distribution parameters are estimated individually which are ultimately combined using logistic regression approach.

3.Methodology

Our aim is to calculate the highest bid price of a given advertisement such that the auction by that bidder is one, keeping in mind not to exhaust the budget suggested.

To solve this, certain terms have to be kept in mind such as total cost and total profit of a campaign

Total Cost = Sum of {Bidding Price * Probability of Bid Winning * Probability of Campaign Being Selected} for all campaigns

Total Profit = Sum of {(Conversion Rate * Payoff Ratio - Bidding Price) * Probability of Bid Winning * Probability of Campaign being selected} for all campaigns

For single campaign, the ‘Probability of Campaign Being Selected ‘will be 1. So the bid price is found on the basis of total cost and the profit generated by an entire campaign. Depending upon the algorithm used by the particular campaign, the price is deciphered.

3.1. Single Campaign

In single campaign we assume all the bids received in a time interval T, which are represented with the help of a feature vector. The meta-bidder calculates a bidding price using a bidding price with the aim of net profit maximization over the time period. The data is initially a 10 column relational table which holds various parameters of the data set such as the timestamp, region, IP, url, url id, city code etc. The data is then pre-processed to reduce the data to keep only the data columns which will be useful. Only 3 columns are required and they are the number of clicks, market price and conversion rate. The probability of selecting a particular campaign in the single campaign procedure is 1.

We want to maximise the target that is the total profit achieved, so the bidding function is estimated and we try to estimate the bidding value given the following constraints:

E[C] <= B (3.1)

The total expected cost spent in time T by the bidder should be less than or equal to the total budget allocated.

Var[R] <= h (3.2)

The variance to measure the total risk should be less than or equal to a pre-defined upper limit h.

0 <= v <=1 (3.3)

The probability of selection of campaign should be between 0 and 1. In single campaign this value is fixed at 1.

$V^T 1 = 1$ (3.4)

The bidding value is calculated using the derivative of the expected value of bidding function and equating it to zero subject to the constraints mentioned above.

$$E[R(v,b(\Theta,r))]=T\sum_{i=1}^{M} \mathrm{v_i} \int \left(\Theta \mathrm{r_i} - \mathrm{b}\ (\Theta, \mathrm{r_i})\right) \mathrm{w}\left(\mathrm{b}\ (\Theta, \mathrm{r_i})\right) \mathrm{p}_\Theta^{\mathrm{i}}\ (\Theta) \mathrm{d}\Theta \qquad (3.5)$$

- We take Lagrangian of the bidding function: L(b(Ɵ,r),λ)
- Take functional derivative of the above equation with respect to bidding function b(Ɵ,r)
- For different winning functions we can obtain optimal values of bidding function.

$$\left(\frac{\Theta \mathrm{r_i}}{1+\lambda} - b(\Theta, \mathrm{r_i})\right)\frac{\partial w\ (b\ (\Theta, \mathrm{r_i}))}{\partial b\ (\Theta, \mathrm{r_i}\ )} = \mathrm{w}(\mathrm{b}(\Theta\ , \mathrm{r_i})) \qquad (3.6)$$

3.2Multiple Campaigns

Here we assume that there are M campaigns for which the bidder can bid for in time interval T. Basically the procedure is decided in 2 steps:

1. E-step – We fix the value of bidding function and calculate the optimal value of probability of selection of campaign subject to the constraints:
    - Var[R] <= h (3.7)
    - 0 <= v <=1 (3.8)

- $V^t 1 = 1$ (3.9)

2. M-step – We fix the campaign selection probably and estimate the optimal value of bidding function with the aim of profit maximization subject to the budget constraint:

    - E[C] <= B (3.10)

We repeat these steps iteratively until the results converge, all the constraints are satisfied and the target is maximized. In expectation maximization algorithm the campaign selection act as a latent factor which helps to calculate the optimal value of bidding function with all constraints satisfied. The M-step is same as the one explained in single campaign; in addition to that we also have to calculate the campaign selection probability to evaluate where to reallocate the volume for each campaign. We use a new parameter γ called as net profit margin. The net profit margin is the ratio of the net profit of the advertising, either from one campaign or a set of them (meta-bidder), divided by the advertising cost during the corresponding period.

$$\gamma = R/C = \text{ROI} - 1 \quad (3.11)$$

Where, ROI is return on investment hence γ is a random variable with variance and expectation. To solve this problem, we use portfolio based risk management methods as described below:

1. Single Campaign: If there is only one single campaign then we calculate the expectation and variance of the net profit margin γ_i for each campaign i. Both $\mu_i(b)$ and σ^2_i (b) can be estimated from MCMC methods:

    (i) Repeat N times on sampling T bid requests from the training data and calculate R_i ($v_{i=1}$; b) and C_i ($v_{i=1}$; b), then

    (ii) Calculate the expectation and variance using these N observations of $R_i(v_{i=1}; b)$ and $C_i(v_{i=1}; b)$

$$\mu_i\,(b) = E\,[\gamma_i] \quad (3.12)$$

$$\sigma^2_i\,(b) = E\left[\frac{R_i\,(v_{i=1}, b)^2}{C_i\,(v_{i=1}, b)^2}\right] - E\left[\frac{R_i\,(v_{i=1}, b)}{C_i\,(v_{i=1}, b)}\right]^2 \quad (3.13)$$

These are the equations used to calculate the mean and variance of the expected value of net profit margin γ.

2. Campaign Portfolio: If we have M campaigns then rather than a single value of mean we would have a vector of expected net profit margins and a covariance matrix for the net profit margins of M campaign where each entry is the correlation factor between campaign i and j, which can be calculated by routine given the net profit margin time series of the two campaigns i and j. This probabilistic campaign combination is called as campaign portfolio, the mean and variance in this case is defined as:

$$\mu_{p}(v, b) = v^{T} \mu(b) \qquad (3.14)$$

$$\sigma 2_{p}(v, b) = v^{T} \Sigma(b)\, v \qquad (3.15)$$

For simplicity, we assume that the net profit margin distribution does not change much w.r.t. the auction volume allocated to the campaign during a short period.

3. Campaign Portfolio Optimization: The E- step can be re written with the help of Lagrangian as:

$$\text{Max } v^{T}\mu(b) - \alpha\, v^{T}\Sigma(b)\, v \qquad (3.16)$$

$$\text{s. t. } v^{T}1 = 1, \quad 0 <= v <= 1$$

Where, the Lagrangian multipliers α acts as a risk-averse parameter to balance the expected net profit margin and its variance. This optimization framework is widely used as portfolio optimization. When the risk, i.e., the variance of the net profit margin, is not considered, α is set as 0. Then the campaign i with the highest $\mu_i(b)$ will be always selected, i.e., $v_i = 1$, while $v_j = 0$ for all other campaigns j. For E-step, the computationally costly parts are the MCMC methods for evaluating the margin of M individual campaign, the complexity is O (MNT) and the campaign correlation calculation which is, O (M^2NT). The downside risk that we have calculated in the project is based on Post Modern Portfolio Theory which has not been used in the previous research paper to the best of our knowledge.

4. Results and Discussion

4.1. Single campaign portfolio result

The following tables show a comparison of the profit for different algorithms (const, rand, truth, linear, ortb, sam1, sam1c, sam2, sam2c) when the budget proportion is 1/16 and pay-off ratio is 0.2

Campaign: 1458

For this particular campaign maximum profit is generated when sam2 algorithm is used.

```
$ bash single_campaign_arbitrage_demo.sh
import the parameters setting from config file
single campaign arbitrage driver
test campaign: ['1458']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
start to train/test
prop    alpha   algo   profit cnvs   bids   imps    budget      cost rratio   para     up
  16  0.00100  const    -2.96    0  80000    740  339072.7    2958.0    0.2    5.0 0.0000
  16  0.00100   rand     0.00    0  80000      0  339072.7       0.0    0.2    5.0 0.0000
  16  0.00100  truth    44.72    3  80000    601  339072.7   11862.0    0.2    1.0 0.0000
  16  0.00100    lin   -93.89   13  59404  10843  339072.7  339097.0    0.2   75.0 0.0000
  16  0.00100   ortb  -131.61   11  55468  12556  339072.7  339091.0    0.2  290.0 0.0000
  16  0.00100   sam1    45.77    6  80000   3148  339072.7   67405.0    0.2   25.0 0.0000
  16  0.00100  sam1c    59.14    6  80000   1303  339072.7   54032.0    0.2   25.0 8.2433
  16  0.00100   sam2    63.92    7  80000   4146  339072.7   68116.0    0.2    7.1 0.0000
  16  0.00100  sam2c    60.63    7  80000   4146  339072.7   71408.9    0.2    7.1 0.7942
```

Campaign: 2259

For this particular campaign maximum profit is generated when sam1 algorithm is used. No particular algorithm generates better results for every data set.

```
$ python ../python/driver_single_campaign_demo.py 2259 0.2
import the parameters setting from config file
single campaign arbitrage driver
test campaign: ['2259']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
start to train/test
prop    alpha   algo   profit cnvs   bids   imps    budget      cost rratio   para     up
  16  0.00100  const    35.35    5  80000  16187  440672.6  212350.0    0.2   23.0 0.0000
  16  0.00100   rand    -0.40    0  80000    197  440672.6     396.0    0.2    5.0 0.0000
  16  0.00100  truth    65.32    4  80000   8920  440672.6  132844.0    0.2    1.0 0.0000
  16  0.00100    lin    89.55    8  80000  13883  440672.6  306771.0    0.2   30.0 0.0000
  16  0.00100   ortb    -2.69    8  80000  18565  440672.6  399011.0    0.2  110.0 0.0000
  16  0.00100   sam1   111.34    8  80000  13322  440672.6  284979.0    0.2   29.0 0.0000
  16  0.00100  sam1c    52.70    7  80000  12629  440672.6  294076.4    0.2   29.0 1.9456
  16  0.00100   sam2    27.61    6  80000  14908  440672.6  269629.0    0.2    4.3 0.0000
  16  0.00100  sam2c    23.38    6  80000  14079  440672.6  273862.2    0.2    4.3 1.6672
```

4.2. Multiple Campaign Portfolio Results

This test includes 3 campaigns with id: 1458, 2259 and 2261. The budget proportion is 1/16 and pay-off ratio is 0.2. We have compared profits for 2 alpha values: 0.1 and 2.

```
$ bash multiple_campaign_arbitrage_demo.sh
import the parameters setting from config file
multiple campaign arbitrage driver
test campaign: ['1458', '2259', '2261']
test algorithms:['const', 'rand', 'truth', 'lin', 'ortb', 'sam1', 'sam2']
payoff ratio:0.2
read in train data...
read in test data...
average campaign original cost test: 6567701.0
test volume: 80000
prop    alpha    algo    profit  cnvs    bids    imps    budget      cost  rratio   para      up
  16  0.10000   const     -4.26     0   80000    1780  410481.3    4264.0     0.2    5.0  0.0000
  16  0.10000    rand    -39.98     1   80000   10304  410481.3  106134.0     0.2   29.0  0.0000
  16  0.10000   truth     44.72     3   80000     601  410481.3   11862.0     0.2    1.0  0.0000
  16  0.10000     lin   -184.14    12   40581   10983  410481.3  410491.0     0.2  105.0  0.0000
  16  0.10000    ortb   -184.16    12   60043   14348  410481.3  410511.0     0.2  290.0  0.0000
  16  0.10000    sam1     45.77     6   80000    3148  410481.3   67405.0     0.2   25.0  0.0000
  16  0.10000   sam1c     49.06     6   80000    2202  410481.3   64108.8     0.2   25.0  3.9500
  16  0.10000    sam2     66.64     7   80000    3905  410481.3   65391.0     0.2    6.5  0.0000
  16  0.10000   sam2c     50.21     6   80000    3100  410481.3   62967.4     0.2    6.5  2.6962
  16  2.00000   const     -3.79     0   80000    1552  410481.3    3787.0     0.2    5.0  0.0000
  16  2.00000    rand     -0.48     0   80000     310  410481.3     477.0     0.2    5.0  0.0000
  16  2.00000   truth     12.26     2   80000    4573  410481.3   72756.0     0.2    1.0  0.0000
  16  2.00000     lin   -184.14    12   40581   10983  410481.3  410491.0     0.2  105.0  0.0000
  16  2.00000    ortb   -184.16    12   60043   14348  410481.3  410511.0     0.2  290.0  0.0000
  16  2.00000    sam1     45.77     6   80000    3148  410481.3   67405.0     0.2   25.0  0.0000
  16  2.00000   sam1c     49.06     6   80000    2202  410481.3   64108.8     0.2   25.0  3.9500
  16  2.00000    sam2     45.56     7   80000    4758  410481.3   86471.0     0.2    6.5  0.0000
  16  2.00000   sam2c     28.05     6   80000    3937  410481.3   85128.1     0.2    6.5  2.6962
```

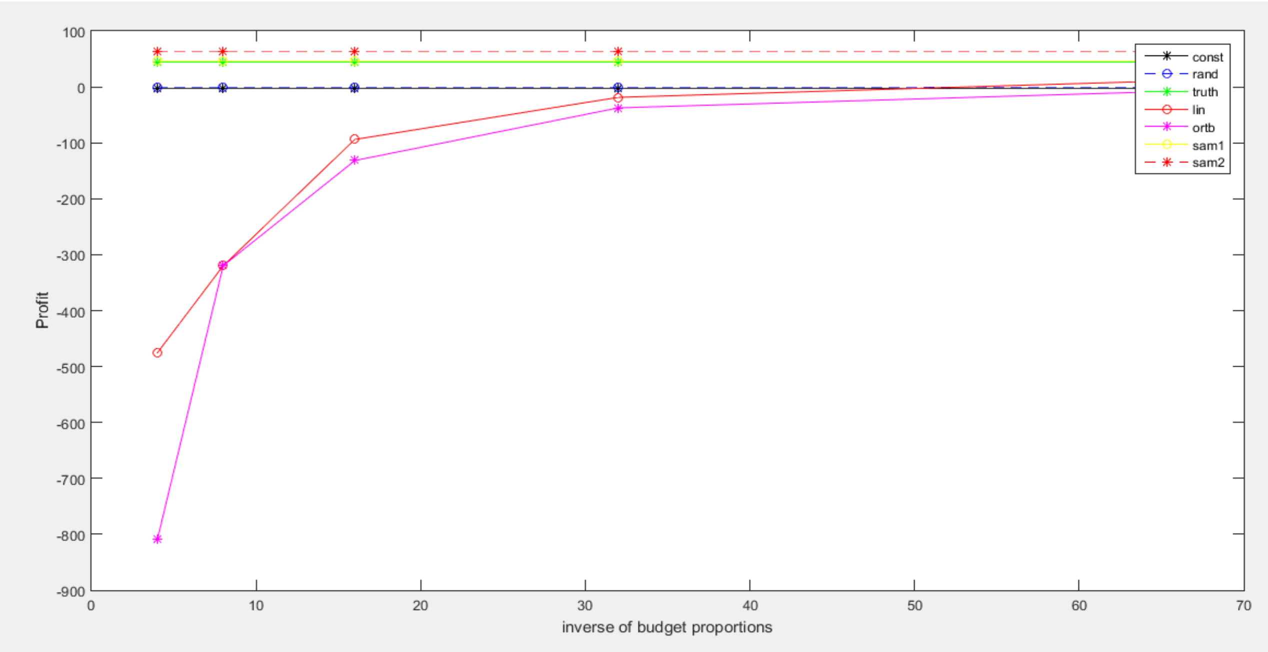
Table 4.1 Profit/Risk for different proportions of budget for SAM1 Algorithm

S.NO	BUDGET PROPORTIONS	PROFIT/RISK
1.	1/4	24.90
2.	1/8	24.90
3.	1/16	24.90
4.	1/32	24.90
5.	1/64	13.51

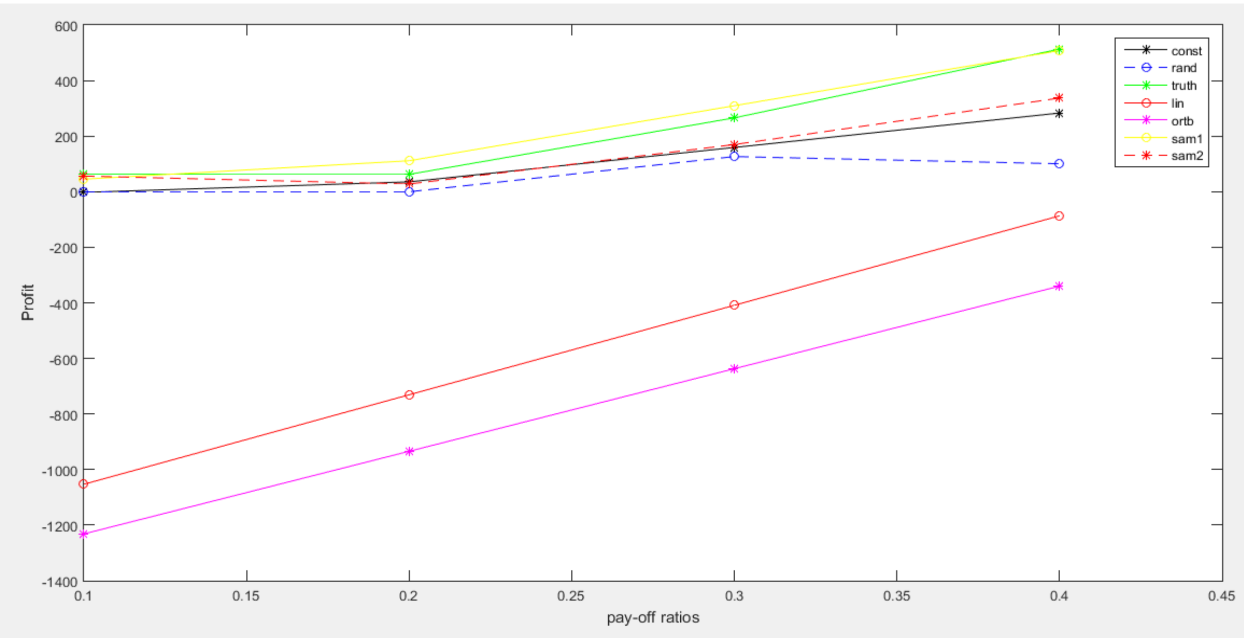
Table 4.2 Profit/Risk for different proportions of budget for SAM2 Algorithm

S.NO	BUDGET PROPORTIONS	PROFIT/RISK
1.	1/4	24.82
2.	1/8	24.82
3.	1/16	24.82
4.	1/32	26.52
5.	1/64	11.11

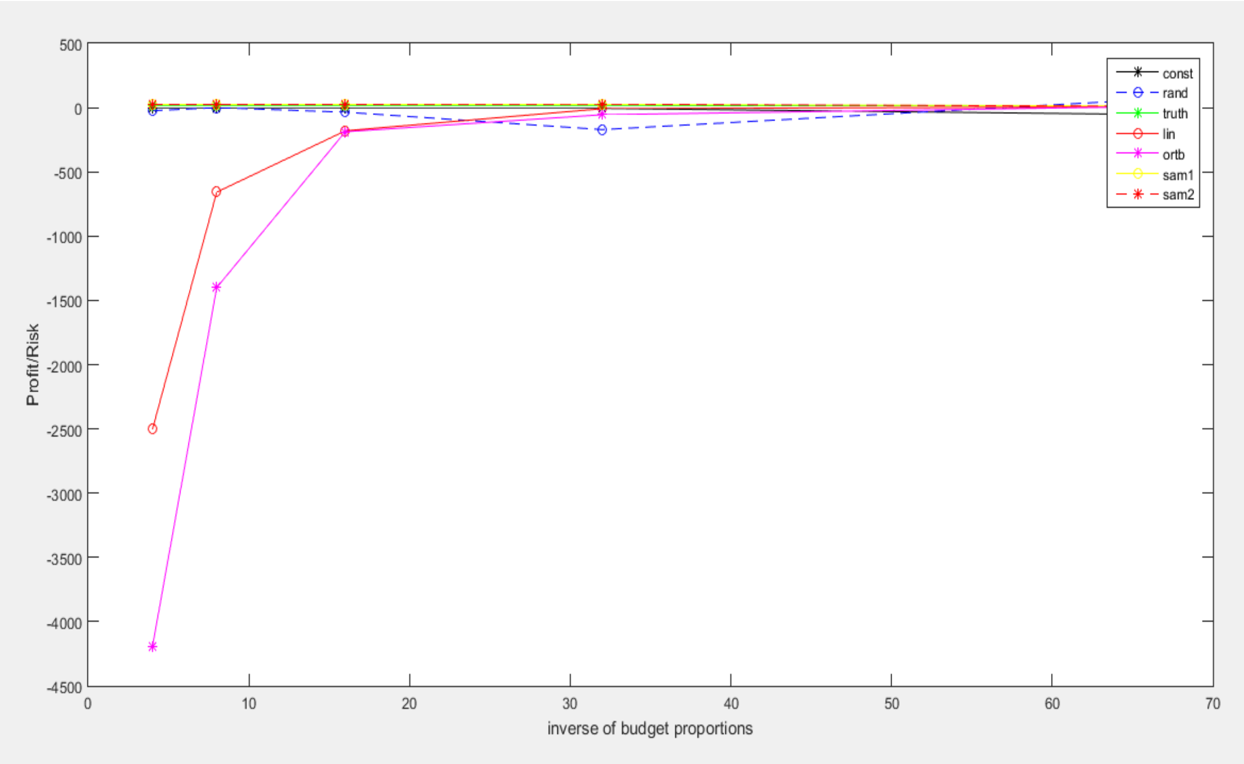
PROFIT FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT BUDGET PROPORTIONS



PROFIT FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT PAY-OFF RATIOS



PROFIT/RISK FOR DIFFERENT ALGORITHMS AS A FUNCTION OF DIFFERENT BUDGET PROPORTIONS



5. Conclusion and Future Scope

In our paper, we have demonstrated and tried to optimize the bidding strategy from the advertisers' side or from the DSP point of view in Real Time Bidding. For the basic single campaign approach, we have optimized the bidding strategy given the constraint of the auction budget by maximizing the profit by maximizing the effective costs per click or the conversion rate. In the more advanced approach that is dealing with multiple campaigns in addition to maximizing the utility we also try to allocate the budget in such a way so as to avoid exhausting the whole budget and investing a larger proportion in the campaign that will lead to a greater revenue. The whole process is based on Modern-Portfolio Theory which takes the advantage of multiple campaigns serves in the DSP. However, for a greater accuracy in the real world (in the process of real time bidding), the downside risk involved in real time bidding is calculated precisely by the Post-modern portfolio theory also proposed in this project.

The concept of Real time bidding came into picture in 2009 and is still an emerging technology which does not only have scope for advertisers and publishers but also possess a lot of scope for improvements. Thus in the future work we would try to find algorithms that maximizes the profit much more and simultaneously reduces the risk as much as possible. We would also try to incorporate the feedback mechanism that is put in action as quickly as possible. This involves updating the bidding price in accordance to the risk found in the previous investment. We would try to make this task iterative and sequential in nature.

6.References

S. a. K. D. Adikari, "Real time bidding in online digital advertisement," in *International Conference on Design Science Research in Information Systems*, 2015.

D. Bloomfield, "High-speed ad traders profit by arbitraging your eyeballs," https://www.bloomberg.com/news/articles/2014-11-07/high-speed-ad-traders-profit-by-arbitraging-your-eyeballs, 2015.

R. P. a. S. V. McAfee, "An overview of practical exchange design," *Current Science(Bangalore),* vol. 109, no. 9, pp. 1056-1063, 2012.

S. Y. J. W. X. S. Weinan Zhang, "Real-time bidding benchmarking with ipinyou dataset," arXiv preprint, 2014.

L. P. Z. L. X. S. Hairen Liao, "iPinYou global rtb bidding algorithm competition dataset," in *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. ACM*, 2014.

B. D. R. H. O. S. T. R. a. F. P. Claudia Perlich, "Bid optimizing and inventory scoring in targeted online advertising," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, China, 2012.

B. O. A. D. a. W. L. Kuang-chih Lee, "Estimating conversion rate in display advertising from past erformance data," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, Beijing, China, 2012.