AdX Writeup

1 Theoretical

1.1 Agent and Game Model

The first assumption for our agent model is that they bid uniformly at random between 0 and 1 on segments that satisfy their campaign's requirement. This lets us calculate the expected max bid for any given segment once we know the expected number of people that are interested in the segment. To calculate the expected max bid, first let X_i be a random variable uniformly distributed between 0 and 1. Assume now that these X_i, the bids for each bidder interested in the segment, are i.i.d. Given we have n bidders on the segment, now assume without loss of generality that X_n is the maximum. We now have that:

$$\mathrm{P}(X_n \leq x) = \mathrm{P}(X_i \leq x \,\forall i)$$

This holds because we know that X_n is the maximum so every other X_i must also be less than x. Now since the X_i are i.i.d. we have:

$$\mathrm{P}(X_i \leq x \,\forall i) = \prod_{i=1}^{n} \mathrm{P}(X_i \leq x) = x^n$$

This is our CDF for X_n which means our PDF is the derivative which is nx^{n-1}. To find the expectation of X_n we can then use:

$$\mathbb{E}[X_n] = \int_0^1 x\,\mathrm{PDF}(X_n)\,dx = \int_0^1 nx^n\,dx = \frac{n}{n+1}$$

Since our agent's goal is to win the segments that satisfy their campaigns as cheaply as possible, we can compare the expected max bids between all the segments that satisfy the campaign and bid primarily on the one with the lowest expected max bid. Because the ad auctions are second price auctions, this expected max bid of the other players in the auction will be our final price that we pay.

Another critical component of this puzzle is estimating n itself – the expected number of users that will be bidding on any given segment. We break this down theoretically into two components: understanding the campaigns that assigned to each user with probability $\min(1, Q)$ and the campaigns that are put up for auction each round.

For the campaigns that are assigned to each user with probability $\min(1, Q)$, we first must approximate Q. Our assumption is that this follows an exponential decay function of the day of the campaign similar to:

$$f(x) = \boldsymbol{e}^{-\frac{x}{10}}$$

This is based on the reasoning that over time people will fail to fulfill their campaigns in part, or in full, and their quality score will suffer. Given this estimation, we know that for the 10 days of the auction, the probability a user is assigned an auction will always be Q since the above function is always less than or equal to 1. For any given campaign we know that the market segment associated with it is chosen uniformly at random amongst all segments with at least two factors (it can't simply be young or male, for example). There are 20 such segments, thus the probability that any given one is chosen is 5%. Segments with two factors, however, can chose between either three factor segment and still satisfy their requirement, thus we associate this 5% probability with both of these segments.

The next type of campaigns are those that are auctioned off. We make the assumption that every campaign that is auctioned off will sell. We then know for certain that someone will be interested in the associated market segments with that campaign. We associate a 100% probability with those segments.

When calculating the expected number of people interested in any given segment, we sum the list of associated probabilities with the given segment. This works because each of these associated probabilities can be treated as the probabilities of indicator random variables and we assume that each are i.i.d. This lets us use the linearity of expectation to sum up these probabilities and arrive at $\mathbb{E}[n]$. This then allows us to calculate the expected max bid of a segment given the math above.

1.2 Optimal Segment Bidding Strategy

Our optimal bidding strategy given the simplified version of the game and our opponents is to use our calculated expected max bids. For segments with two attributes, we should only bid on the three attribute subsegments. This way, we can choose the subsegment that is "cheapest". For segments with three attributes, we only have one segment that can fulfill our campaign so we bid on just this segment. For each segment that applies to one of our campaign's reach, we bid the minimum of *just above* the expected max bid and and $\frac{B}{R}$, the campaign's budget per reach. This way, in expectation, we are winning all segments we bid on and we also don't spend more than our budget to get to an effective reach of one. We set the limit of each segment to be $b(B - P)$ where b is the bid for the segment, B is the budget, and P is the money already spent on the campaign.

The next part of our segment bidding strategy prevents our model from bidding on the same segment twice (for two different campaigns) and effectively erasing our profit since we would have bid the same amount on the segment for both campaigns. To calculate the priority we weight together the number of attributes in the segment, the current reach remaining of the campaign, and the budget of the campaign. Intuitively, if a campaign is specifically associated with a three attribute segment (like male, young, and high income) they have no other option then to bid on this specific campaign. It would therefore be a losing strategy to reduce our bid for this segment for the given campaign. For effective reach remaining, we want to prioritize campaigns who still have the most remaining reach unfulfilled. This protects our quality score from being hurt from not finishing our campaigns. Finally, we want to prioritize campaigns with budget remaining, otherwise even winning the campaign could drive negative profit as we overspend on the campaign. Weighting these together we arrive at our priority score for each bid.

Once our bids are sorted by priority for each segment, we significantly shade all but the highest priority bids. This increases the likelihood that our final price will be extremely low and we will give ourselves a large amount of profit.

1.3 Optimal Campaign Bidding Strategy

The optimal bidding strategy is to try to get as many campaigns that have overlapping segments as possible. This includes campaigns that have segments equal to segments of our current campaigns, segments that are subsets of the segments of our campaigns, and segments that are supersets of the segments in our campaigns. This is because if we can win all of the campaigns that overlap with the users that we are bidding on, we can effectively set the price for our impressions in that segment. On the other hand, if an adversary were to win these campaigns, we would have to outbid them. An optimal strategy will bid more aggressively on campaigns that have segments that are subsets or supersets of segments that we have in our campaigns and will bid even more aggressively on campaigns that have segments equal to those in our campaigns. Additionally, campaigns up for auction that are “overlap” with more than one of our campaigns’ segments will be even more appealing. The function that decides exactly how much to decrease our campaign bids by is not apparent at this time but will be optimized through model testing.

2 Practical

2.1 Evaluation Tools

In order to better evaluate our model, we collect all the data printed to the terminal. For each day of every game and for every player, we track cumulative profit, quality score, current campaigns and their statistics. We do this by redirecting the Java print statements to an output file and parsing the file. Below we provide the graphs and statistics we used to evaluate and improve our strategies and optimize hyperparameters.

	AlexSean	**bot_1**	**bot_2**	**bot_3**
Profit	2970.15	1492.08	1473.59	1477.43
Num Campaigns	6457	5030	4986	5006
Weight Positive	80	80	80	80
Weight Negative	20	20	20	20
% Profitable	62.1	42	40.4	40.9
% Zero	22.5	41.5	43.4	42.7
% Negative	15.4	16.5	16.2	16.3
Num Profitable	4011	2115	2015	2048
Num Zero	1453	2085	2162	2140
Num Negative	993	830	809	818
Eff. R.	0.484	0.19	0.179	0.18
Total Imp.	6262862	3443134	3249837	3324711
$/Imp.	0.554	0.525	0.523	0.529

Table 1: Performance vs. Tier 1 Bots

2.2 Segment Bidding Strategy

Our segment bidding strategy closely follows our theoretical optimal segment bidding strategy desribed above. The exact bid b for each three attribute segment is given by:

$$b = 1.3 \cdot b_E \cdot \delta$$

where b_E is the expected max bid, $\delta = \sqrt{1.38442^2 - \mathrm{ER}(x, R)^2}$, 1.38442 is the maximum expected reach, and $\mathrm{ER}(x, R)$ is the campaign's current expected reach given current impressions x and reach R. Then, we apply a floor and ceiling:

$$b = \min(\max(b, 0.5), B/R)$$

We noticed that using the theoretically optimal bidding strategy described above, our model had far too many campaigns with zero profit and that for these campaigns, we had zero impressions. To counteract this we decided to increase our bids. We include the δ term to shade bidding by how far we are from the maximum effective reach. This term was chosen because it contains the points $(x = 0, 0)$ and $(x = 1.38442, 1)$ while exaggerating the difference between the current effective reach and the maximum effective reach. Once we get close enough to the maximum effective reach, our returns for more impressions decreases so we decrease our bids. But, if we are at 20% or 50% of the maximum effective reach, we increase our bids by a factor of 1.36 and 1.20 respectively to try to take advantage of the large slope in the middle of the sigmoidal effective reach function. We also added the coefficient of 1.3 and the bid "floor" of 0.5 as hyperparameters that we tuned using trial and error. We set each segment's limit to be: $\dot{b}(B - P)$ in accordance to our limit theory which seemed to work well since our negative profit campaigns were rarely from getting too many impressions.

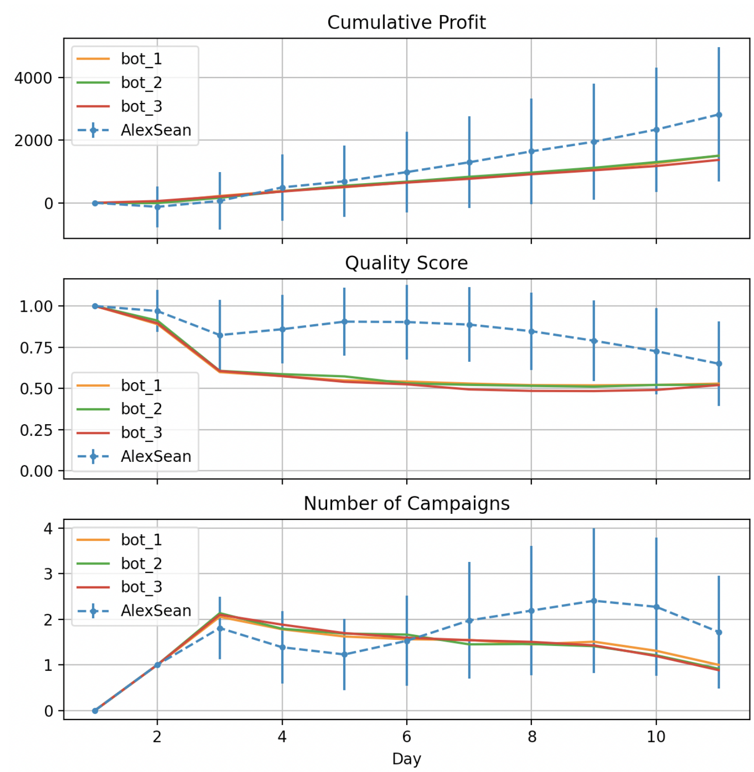

Figure 1: Average $(\pm\sigma)$ of Several Metrics over Time

From here, we go through our campaigns and see if we bid on any segment twice for reasons described above in the theory section. By trial and error, we found a reasonable measure for priority to be:

$$\text{campaign segment num attributes} \cdot \delta \cdot \frac{B}{t}$$

This made sense because if a campaign segment has three attributes, this is the only segment we can get impressions from. $\delta \cdot \frac{B}{t}$ tells us approximately how much more revenue is left in this campaign and we divide by t to level the playing field between campaigns of different durations. With more time, we could have made this term depend on the probability each campaign would be filled considering the other segments a campaign bids on. After sorting the overlapping bids, we shade bids and their limits by $\frac{1}{(i+1)^2}$ where i is the campaign's index in the list sorted by the priority heuristic described above. This keeps the more prioritized campaign's bid the same while drastically decreasing the bid size with each subsequent priority. Finally, we set the campaign spending limits to be $\frac{B-P}{t_{left}}$ because we found that we decreased our costs by not spending too much on one day.

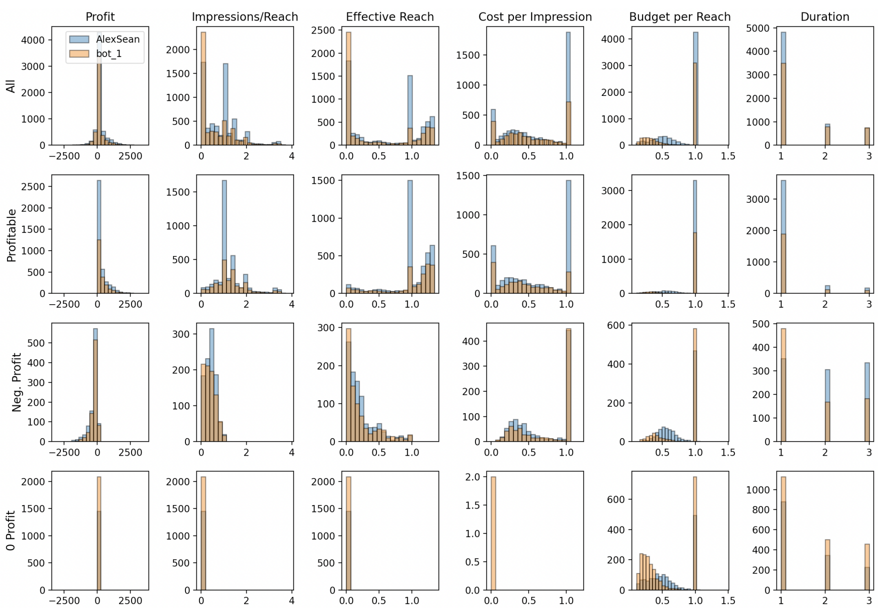

Figure 2: Histograms for (all, positive, negative, zero) Profit Campaigns

2.3 Campaign Bidding Strategy

Our campaign bidding strategy followed our theoretical strategy very closely and we start with the expected max bid of segments that have three attributes and the mean of the subsegments of segments that have two attributes. Next, if the segment is a two attribute segment, we chose a discount factor of 0.75 in accordance to our theory and by testing different values and evaluating the model's performance. We then try to get the campaigns that have overlapping segments with the segments of our current campaigns. We settled on a discount factor of 0.8 for campaigns with segments equal to the segments of our campaigns and a discount factor of 0.95 on segments that are subsegments of segments in our campaigns or vice versa. In a more detailed model, this discount would also consider the durations of each of the overlapping campaigns. For both segment and campaign bidding, an assumption we make is that the functions and hyperparameters that we chose could reasonably optimize our performance. We chose a lot of coefficients that intuitively made sense to us but whether or not they were the best choice remains unclear as our performance heavily depends on the strategies of our opposition.

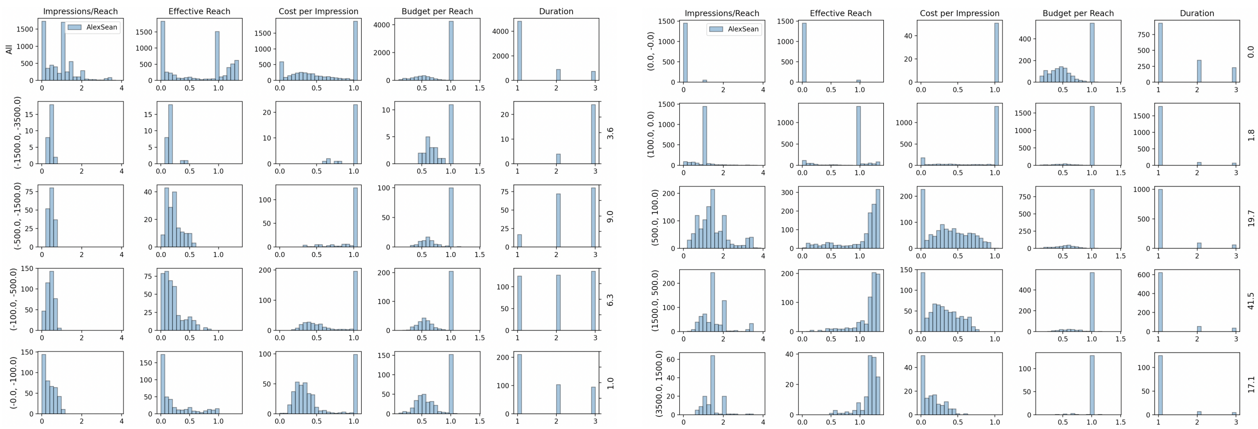

Figure 3: Histograms for More Specific Profit Segment Campaigns