Bid-war: Human or Robot?
June 2, 2015

Charles Packer
cpacker@ucsd.edu

William Huang
wihuang@ucsd.edu

Abstract

Online auctions have become an increasingly important aspect of e-commerce. However, for consumers, the chances of landing a winning bid in online auctions has become increasingly more difficult with the abundance of "bidding robots". Bidding robots such as "BidRobot" and "Auction Sniper" are pieces of software that are configured by the user to follow any number of auctions on different auction sites simultaneously, bidding in place of the user according to predefined settings and preferences. Humans are not able to attend to and monitor auctions with the same capacity as a computer, which can make complex bidding decisions in split-second time and can follow an auction with nonstop, undivided attention. Thus bidding robots gain a significant competitive advantage over their human counterparts, posing an interesting problem for auction sites that wish to level the playing field by identifying and banning bid bots. In this paper, we seek discover methods to predict whether certain bidders are in fact humans or bidding robots.

1 Introduction

For the purpose of predicting whether bids are made by humans or bidding robots, we will be using the dataset released for a Kaggle Competition (Facebook Recruiting IV: Human or Robot?)[1] This data is from some online auction platform that "may or may not be affliated with Facebook" [2]. The data is organized into three files: `train.csv` (bidder data with labels), `test.csv` (bidder data without the labels meant to be tested against Kaggle's hidden set with labels), and `bids.csv` (individual bid data from 7.6 million bids). For the purpose of this paper, we will be using the labelled bidder data and bid data to make and test our predictions.

2 Dataset Analysis

2.1 Data Fields

For the bidder dataset (`train.csv`):

- `bidder_id`: Unique identifier of a bidder.
- `payment_account`: Payment account associated with a bidder. These are obfuscated to protect privacy.
- `address`: Mailing address of a bidder. These are obfuscated to protect privacy.
- `outcome`: Label of a bidder indicating whether or not it is a robot. Value 1.0 indicates a robot, where value 0.0 indicates human.

For the bid dataset (`bids.csv`):

- `bid_id`: unique id for this bid
- `bidder_id`: Unique identifier of a bidder.
- `auction`: Unique identifier of an auction
- `merchandise`: The category of the auction site campaign
- `device`: Phone model of a visitor
- `time`: Time that the bid is made (transformed to protect privacy).
- `country`: The country that the IP belongs to
- `ip`: IP address of a bidder (obfuscated to protect privacy).
- `url`: url where the bidder was referred from

2.2 Exploratory Analysis

We begin by examining the data to seek out important and key behaviors that differentiate human and bot bidders. In order to better direct our initial investigation, we surveyed a variety of research to on the topic of bot bidders and their traits to help us decide what are potential features and facts that we can use to improve precision and accuracy for our predictive model.

There are 6614 unique bidders and 15051 unique auctions in the dataset. Among the `bidder_ids` referred to in the training data, only 29 out of 2013 (29 humans, 0 bots) are unaccounted for in the bids data, meaning that the vast majority of labelled bidders have accompanying bid data (in addition to their address and payment account). This means the two datasets can for the most part be treated as one large set.

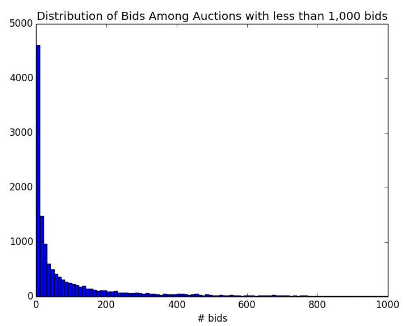


Figure 1: Bids Per Auction (less than 1000 total bids)

The mean number of bids per `bidder_id` is 1157.60, with over 91.50% of bidders having made less than 1000 bids and 86.00% having made less than 500 bids (see figure 1). The mean number of bids per auction is 508.70, however, among humans exclusively that number becomes 6.44, and for bots it becomes 23.15. This difference is apparent in the distribution shown in figure 3. A similar difference in distribution occurs for total number of bids made by humans versus bots, as shown in figure 2. Figure 2 and 3 also demonstrates that the average in both cases is raised by outliers (bots) with a high number of total bids (greater than 4000) and a high number of average bids per auction (greater than 40). This makes sense, since a bot that is constantly watching the auction may be more likely to bid frequently (in the same auction) in order to outbid competitors. One might also argue that an experienced user (that has a high number of total bids) may be more likely to employ the use of bots for the competitive advantage they provide (conversely, those who do not use bots may be those who participate less frequently as a casual bidder, or possibly quit the marketplace in frustration having only made a few number of bids).

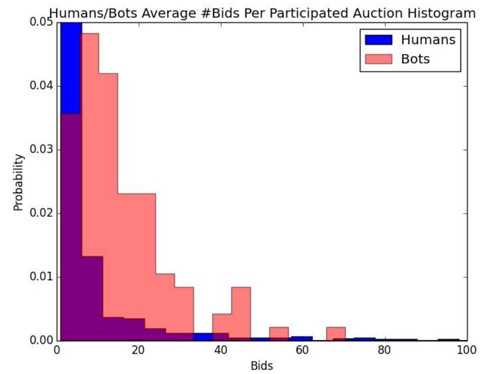


Figure 2: Human and Bot Average Bids Per Auction

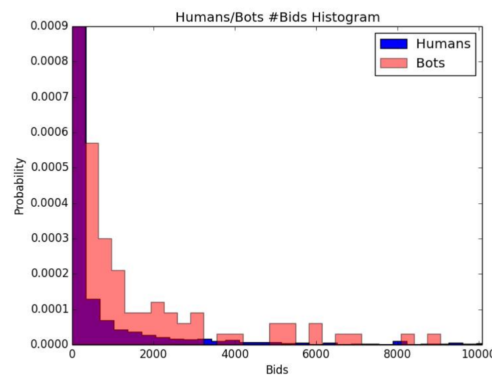


Figure 3: Human and Bot Average Total Bids Made

From our analysis we can also identify features of the data that most likely will not be of use in differentiating between bots and humans. In figures 4 and 5, we see that the distribution of bids and auctions among categories (roughly equivalent to popularity) is about the same, with no outstanding differences. The distribution of bot activity among categories (figure 6) for the most part follows this same distribution, and so knowing which categories a bidder is most active in most likely will not be of much use in determining whether or not the bidder is a bot.

Similar to our investigation of categories, we decided to examine the number of known bots in the training set and their country of origin. Potentially this may allow us to use a predictor that may find some bias towards origin of the bidder from this trend. Figure 8 shows the top 100 countries with highest bot activity, and figure 7 shows the distribution among all bids for the top 100 countries in the dataset. After computing the shift in ranking among countries in total bids and bot bids, we found the a negligible difference, and decided not to use country data as part of our initial models.

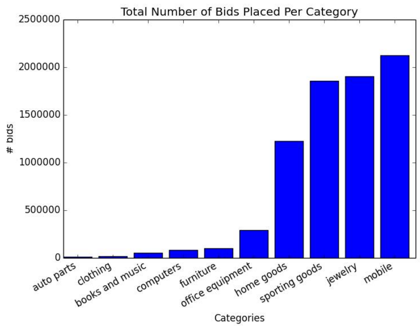


Figure 4: Total Bids Per Category

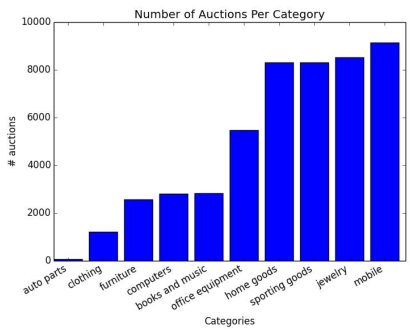


Figure 5: Total Auction Per Category

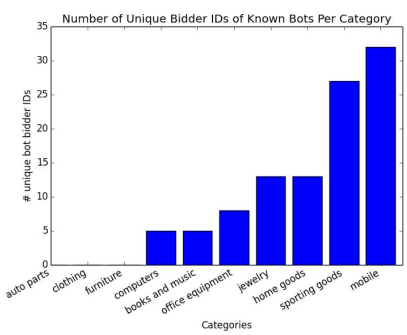


Figure 6: Bot Activity Per Category

With that in mind however, while some discoveries may not seem significant for our predictive task, we do not want to prematurely dismiss any discoveries just yet. For example, the figures showing the difference in distribution among countries for all bids and only bot bids shows a slight difference, which may not prove useful for the initial model, but might prove useful for fine-tuning later on in order to achieve higher accuracy predictions. Thus although the change in distribution may be insignificant, we still take note of such differences. Due to time

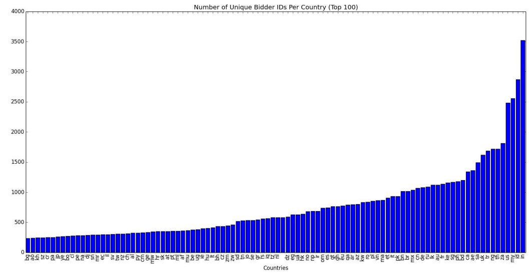


Figure 7: Distribution of All Bids Among Countries (top 100)

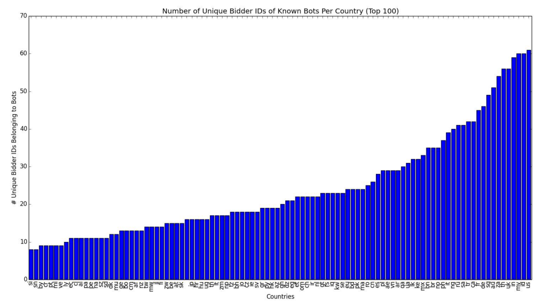


Figure 8: Distribution of Bot Bids Among Countries (top 100)

constraints, other parameters such as `device` and `time` were not investigated, however they provide an interesting opportunity for further research on the topic.

3 Predictive Task Methodology

3.1 Evaluation

The predictive (supervised learning) task proposed is to predict whether a given bid is made by a human or bot, given a training set with known labels. An output of 1.0 from our predictor suggests that the bid is made by a bot, otherwise, an output of 0.0 from our predictor suggests that the bid is made by a human.

3.1.1 Sensitivity and Specificity

Classification accuracy and error rate. [3]

		Label	
		true	false
Prediction	true	true positive	false positive
	false	false negative	true negative

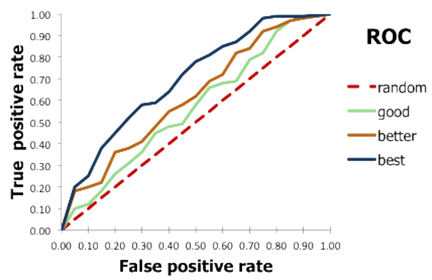
Classification accuracy	= correct predictions / #predictions
	= (TP + TN) / (TP + TN + FP + FN)
Error rate	= incorrect predictions / #predictions
	= (FP + FN) / (TP + TN + FP + FN)

3.1.2 Receiver Operating Characteristic

We used ROC area-under-curve to evaluate our classifiers. ("The curve is created by plotting the true positive rate against the false positive rate at various threshold settings.") [6]

$$TPR = TP/P = TP/(TP+FN)$$

$$FPR = FP/N = FP/(FP+TN)$$



3.1.3 Validation

In terms of validating our classifier, we use a 10-fold cross validation pipeline [5]: the training data is broken into 10 sets, 1 of which is used for testing, and the other 9 for training. The results are then computed from 10 rounds and the mean performance (ROC-AUC) is recorded.

3.2 Relevant Baseline

We provide pseudo-code for a simple baseline below, where the threshold is calculated by $\mu + c$, where μ equals mean bids per auction among all human bidders (a metric chosen as a result of the exploratory analysis), and c is some constant:

```
if we have bid data for the bidder:
    if average bids per auction > threshold:
        prediction = 1.0
    else:
        prediction = 0.0
else:
    prediction = 0.0
```

where `human_mean_bidsPerAuction = 6.44` and `bot_mean_bidsPerAuction = 23.15` (as previously described).

By adjusting c to 20 (approximately the mean bids per auction among all bot bidders), the baseline proved fairly effective. Using the full data, we extracted the followings. The table below is the result from going over all the train data, making a prediction, then looking at the actual prediction and see if it was correct (the equivalent of training error). The first row is the constant c added to μ. The second row is the resulting accuracy defined from the following equation:

$$\sum_{i=1}^{n} PredictedBot \div \sum_{i=1}^{n} ActualBot$$

$c = 1$	$c = 10$
0.86	0.91

3.3 Higher-Dimensional Models

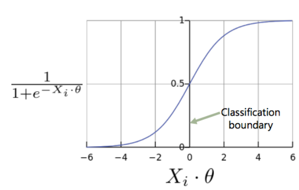
While we observe high accuracy by using the baseline model, we wish to apply a more sophisticated model which can incorporate more features from our data. The task at hand deals with binary labels and therefore we can use predictive methods that are designed for binary classification. Despite the addition of other features, the bids per auction metric however proves to be very powerful in classifying our data, as we will see. As is true in many cases, we found that switching between classifiers rarely provided more than a marginal increase in performance, whereas altering features had the potential to drastically alter the classifier's accuracy.

3.3.1 Logistic Regression

Since we are dealing with binary outputs and the training example(s) can be structured as f(data) = labels, we can use Logistic Regression, which provides strong results with a relatively simple model:

$$y_i = \begin{cases} 1 & \text{if } X_i \cdot \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

sigmoid function: $\sigma(t) = \frac{1}{1+e^{-t}}$



3.3.2 Related Models

In testing, we applied a variety of supervised learning algorithms, such as Support Vector Machines and decision trees (random forest). Other algorithms such as k-Nearest Neighbor (kNN) can also be plausible classifiers

- the downsides however are 2-fold when working on a machine with limited RAM since we are dealing with a large dataset:

- Space (RAM) requirement is high
- Classification is computationally expensive

4 Results and Model Analysis

In terms of feature extraction, the best results were achieved when using only two features: average bids per auction, and total bids per user. Many other feature combinations were attempted, including binning average bids and total bids based on the human-bot distribution graphs generated earlier. Despite these attempts, these two features alone proved to be very powerful in classifying bot versus human bidders, and no significant gains in accuracy were made by incorporating other features such as categories (merchandise) or country.

Regarding the best predictive model, no models tried were able to show meaningful performance boosts over the simpler logistic regressor. Ultimately, the combination of two key features in conjunction with logistic regression performed the best in classifying bot bidders with high accuracy.

Model(s) Results [5]			
Logistic Regression	AdaBoost	SVM	Random Forest
0.864455	0.874700	0.862054	0.763824

5 Related and Future Work

Regarding other related studies, we found a study of Botnet detection to be somewhat similar. In the study, several classifiers were used [4].

- Multinomial Naive Bayes
- linear SVM
- kNN
- Logistic Regression
- Multiboost Adaboost

We believe that Botnet detection can be similar because it involves detecting similar traits in the data to discern between human and non-human activity. Furthermore, the end-goal remains the classify whether a action is made by a human or bot. One such feature may be the IP address (certain IPs may be linked to high bot-traffic). One key difference between detecting botnets and bid bots is that while botnets by name imply certain network (graph) properties, bid bots often run on isolated software on the users own device.

While we were unable to analyze certain aspects of the data due to time constraints, we feel that the following features may prove effective in differentiating between human and bot activity: `time` (bots may have lower average time between consecutive bids on the same auction due to automated auction tracking) and `url` (bots may be referred from an address related to the bidding software used).

6 Conclusion

Bid bots are detrimental for online auction sites. Take for example, bidders who ask "What is the point of bidding when people have bots that will out bid you at the last second?" [1] For serious bidders, the best solution may be to use bots to beat the bots, feeding a positive feedback loop in turn creating a bot-controlled marketplace. Bid bots usually perform with some sort of trend and pattern, precisely the sort that allows them to be so effective in winning auctions (e.g. bidding multiple times at the last second before an auction closes). Using auction data we can discover these trends and patterns, allowing us to predict whether a bid is made by a human or a bot. Although, while it may be difficult to attain full precision and accuracy in classifying bot activity with only the metrics used in our study, we can at least predict with enough accuracy to make meaningful conclusions about our assumptions on human versus non-human behavioral patterns.

References

[1] EBAY. Bot bids. *https://community.ebay.com/t5/Archive-Getting-Started/What-is-the-point-of-bidding-when-people-have-bots-that-will-out/td-p/16843159*.

[2] KAGGLE. Facebook recruiting iv: Human or robot? *https://www.kaggle.com/c/facebook-recruiting-iv-human-or-bot*.

[3] MCAULEY, J. Cse 190: Dahta mining and predictive analytics. *http://cseweb.ucsd.edu/ jmcauley/cse190/*.

[4] NIVARGI, V. Machine learning based botnet detection. *http://cs229.stanford.edu/proj2006/NivargiBhaowalLee-MachineLearningBasedBotnetDetection.pdf*.

[5] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., AND DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12* (2011), 2825–2830.

[6] WIKIPEDIA. Receiver operating characteristic. *http://en.wikipedia.org/wiki/Receiver_operating_characteristic*.