Math642\_Project Plan

Fyona Sun

2020/02/12

**Topic**

Predicting the end-price of online auction

**Objective**

Online auctions are one of the most popular methods to buy and sell items online. It is part of the growing body of eCommerce system and can be further apply to other competitive markets. EBay, Yahoo! Auctions, Amazon Marketplace have started this auction methods a long time ago, recent year there are new sites who focus on creating a second-hand and sustainable environment also adopted this bidding system. The activities of online auction produce a large number of transaction data. The end price prediction results may help sellers optimize the selling price of their items and auction attributes and the bidder can get a good estimation of item with their budget. As a result, the transaction time can be shortened, and the cost can be reduced.

**Problem Description / Need**

Because of the nature of bidding behavior, one big challenge is to capture the competition between bidders. Predicting the auction end-price by categorical data of the item is not enough. Modeling differences in product similarity and their impact on bidders’ choices are also challenging.

When I was doing my research in this area, I found that the existence of buyer-seller networks and the impact of these networks on transaction outcomes also very interesting. I would also include the impact of user networks in this project if possible.

**Dataset Description**

Source: <https://www.kaggle.com/onlineauctions/online-auctions-dataset>

The datasets contain eBay auction information on Swarowski beads.

The dataset contains two different types of data, the bid history and the item features. The bid history is a sequence of bids placed over time, which can be considered as a time series type of data. The auction on eBay are fixed-length auctions with length usually ranging from 1 to 10 days. Therefore, the dataset is stored separately according to how long the auction lasts. The other set of data describes the item to be auctioned, which does not change over time, and it contains categorical and descriptive information.

The bidding history has the structure like:

|  |  |
| --- | --- |
| auctionid | the unique identifier of an auction |
| bid | the proxy bid placed by a bidder |
| bidtime | the time in days at which the bid was placed (from the start of the auction) |
| bidder | username of the bidder |
| bidderrate | bidder's rating |
| openbid | the opening bid (set by the seller) |
| price | the price that the item was sold for (equivalent to the second highest bid + an increment) |

**Analysis Plan**

The Kaggle dataset did not provide the categorical data of the items that’s on auction, which makes it difficult to apply regression models. By looking at the auction dataset, I will be focusing on the price curve representation, which estimate the price process during an ongoing auction.

Then I would explore the auction networks.

Finally I would build a model to predict the end-price of the auction with the information from within a given ongoing auction to forecast its final price.

Final\_Project\_FyonaSun

Fyona Sun

2/10/2020

library(readr)  
auction=read\_csv('./online-auctions-dataset/auction.csv')

## Parsed with column specification:  
## cols(  
## auctionid = col\_double(),  
## bid = col\_double(),  
## bidtime = col\_double(),  
## bidder = col\_character(),  
## bidderrate = col\_double(),  
## openbid = col\_double(),  
## price = col\_double(),  
## item = col\_character(),  
## auction\_type = col\_character()  
## )

head(auction)

## # A tibble: 6 x 9  
## auctionid bid bidtime bidder bidderrate openbid price item   
## <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <chr>  
## 1 1.64e9 175 2.23 schad… 0 99 178. Cart…  
## 2 1.64e9 100 2.60 chuik 0 99 178. Cart…  
## 3 1.64e9 120 2.60 kiwis… 2 99 178. Cart…  
## 4 1.64e9 150 2.60 kiwis… 2 99 178. Cart…  
## 5 1.64e9 178. 2.91 eli.f… 4 99 178. Cart…  
## 6 1.64e9 1 0.356 bfalc… 2 1 355 Cart…  
## # ... with 1 more variable: auction\_type <chr>

dim(auction)

## [1] 10681 9

The data set includes 10681 observations and 9 variables auctionid: unique identifier of an auction bid: the proxy bid placed by a bidder bidtime: the time in days that the bid was placed, from the start of the auction bidder: eBay username of the bidder bidderrate: eBay feedback rating of the bidder openbid: the opening bid set by the seller price: the closing price that the item sold for (equivalent to the second highest bid + an increment) item: auction item auction\_type

swarovski.csv is the seller-buyer network includes 5 variables: Seller Bidder Weight Bidder.Volume Seller.Volume

auction=read\_csv('./online-auctions-dataset/auction.csv')

## Parsed with column specification:  
## cols(  
## auctionid = col\_double(),  
## bid = col\_double(),  
## bidtime = col\_double(),  
## bidder = col\_character(),  
## bidderrate = col\_double(),  
## openbid = col\_double(),  
## price = col\_double(),  
## item = col\_character(),  
## auction\_type = col\_character()  
## )

swarovski=read\_csv('./online-auctions-dataset/swarovski.csv')

## Parsed with column specification:  
## cols(  
## Seller = col\_double(),  
## Bidder = col\_double(),  
## Weight = col\_double(),  
## Bidder.Volume = col\_double(),  
## Seller.Volume = col\_double()  
## )

head(auction)

## # A tibble: 6 x 9  
## auctionid bid bidtime bidder bidderrate openbid price item   
## <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <chr>  
## 1 1.64e9 175 2.23 schad… 0 99 178. Cart…  
## 2 1.64e9 100 2.60 chuik 0 99 178. Cart…  
## 3 1.64e9 120 2.60 kiwis… 2 99 178. Cart…  
## 4 1.64e9 150 2.60 kiwis… 2 99 178. Cart…  
## 5 1.64e9 178. 2.91 eli.f… 4 99 178. Cart…  
## 6 1.64e9 1 0.356 bfalc… 2 1 355 Cart…  
## # ... with 1 more variable: auction\_type <chr>

head(swarovski)

## # A tibble: 6 x 5  
## Seller Bidder Weight Bidder.Volume Seller.Volume  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 332874919 718577508 2 3 547  
## 2 594667804 399983466 5 6 183  
## 3 663070601 655828811 1 4 274  
## 4 309608641 599835541 3 8 3986  
## 5 201729374 693022555 1 2 4681  
## 6 332874919 622536435 2 4 547

dim(auction)

## [1] 10681 9

dim(swarovski)

## [1] 200 5

summary(auction)

## auctionid bid bidtime   
## Min. :1.639e+09 Min. : 0.01 Min. :0.000567   
## 1st Qu.:3.015e+09 1st Qu.: 72.00 1st Qu.:1.949931   
## Median :3.021e+09 Median : 140.00 Median :4.140833   
## Mean :4.136e+09 Mean : 207.59 Mean :3.979628   
## 3rd Qu.:8.212e+09 3rd Qu.: 210.00 3rd Qu.:6.448060   
## Max. :8.216e+09 Max. :5400.00 Max. :6.999990   
##   
## bidder bidderrate openbid price   
## Length:10681 Min. : -4.00 Min. : 0.01 Min. : 26.0   
## Class :character 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 186.5   
## Mode :character Median : 5.00 Median : 4.99 Median : 228.5   
## Mean : 31.94 Mean : 52.25 Mean : 335.0   
## 3rd Qu.: 21.00 3rd Qu.: 50.00 3rd Qu.: 255.0   
## Max. :3140.00 Max. :5000.00 Max. :5400.0   
## NA's :11   
## item auction\_type   
## Length:10681 Length:10681   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##

summary(swarovski)

## Seller Bidder Weight Bidder.Volume   
## Min. : 24273 Min. : 221355 Min. : 1.000 Min. : 1.00   
## 1st Qu.: 94870114 1st Qu.: 80389757 1st Qu.: 1.000 1st Qu.: 4.00   
## Median :201729374 Median :191077522 Median : 2.000 Median : 6.00   
## Mean :198089751 Mean :276467688 Mean : 3.085 Mean : 9.26   
## 3rd Qu.:309608641 3rd Qu.:449791350 3rd Qu.: 3.000 3rd Qu.:11.00   
## Max. :663070601 Max. :723365710 Max. :56.000 Max. :94.00   
## Seller.Volume   
## Min. : 2.0   
## 1st Qu.: 382.5   
## Median :1791.0   
## Mean :1966.3   
## 3rd Qu.:3986.0   
## Max. :4681.0

# hierarchical clustering with complete linkage and Euclidean distance

hcmodel.complete <- hclust(dist(swarovski), method = "complete")  
plot(hcmodel.complete)

A close up of a logo

Description automatically generated

cutree(hcmodel.complete, 3)

## [1] 1 1 1 1 2 1 3 3 2 2 3 3 2 3 3 3 1 3 3 3 3 3 3 3 2 2 3 1 3 3 3 2 1 2 3  
## [36] 3 3 3 3 2 1 1 3 3 2 2 3 3 3 3 3 3 3 2 3 2 3 2 3 2 3 3 3 3 2 3 3 3 3 3  
## [71] 2 3 3 2 1 1 2 3 2 2 3 3 3 2 3 3 3 1 2 1 3 3 3 3 1 1 3 3 2 2 3 3 3 2 3  
## [106] 3 3 3 3 3 2 3 3 3 3 3 3 1 2 2 3 2 2 3 2 3 1 1 1 3 3 3 3 2 3 3 3 3 2 3  
## [141] 2 3 3 2 3 3 2 3 2 3 3 3 3 3 3 2 1 1 2 1 3 3 3 2 1 3 3 3 3 3 3 2 3 2 3  
## [176] 3 2 2 3 3 3 3 1 2 3 2 1 2 3 3 3 2 3 3 3 1 3 3 3 3

Appendices–Linear Regression–KNN–LASSO–RR–PCR–PLS

Math642\_Project Plan Revision

Fyona Sun

2020/02/26

**Objective:**

Forecasting the auction outcome, that is the final price is beneficial to all auction parties. Especially I want to build a model that does not just measure where the price will end up, but how fast the price is moving along to its end price.

Bidders can use price forecasts to make more informed bidding decisions based on their budgets. Since it’s difficult for a bidder to monitor all the active auctions, it might be difficult for a bidder to decide which of these numerous auctions to participate in and to place a bid that does not end up overpaid or failed to win the auction.

Sellers can use price forecasts to determine when to post their listing. There are companies that re-sell materials on eBay could also make use of price forecasting. These stores sell materials on behalf of individuals who do not want to use eBay directly. Ongoing price forecasting would allow these stores to pay their customers for their merchandise before the close of the auction, in effect offering faster liquidation for their customers.

Auction houses can use forecasts for long-term budgeting and planning purposes or even real-time adjustments during the bidding process. Another way to think in the online setting is that a dynamic price scoring would allow auctions to be ranked by lowest expected price, which would in turn allow purchasers to focus their time and energy on just those few auctions that promise the lowest price. Thus a well-functioning forecasting system could be adaptive to accommodate and incorporate change in the dynamic setting.

**Analysis Plan**

I can first build a static model using information known before the start of the auction, such as the opening bid, the auction length, and the seller’s reputation. Although this traditional method of forecasting prices are hard to apply and are not very accurate for online auctions, because they don’t take into account the dramatic changes in auction dynamics. Price changes in online auctions do not happen at a steady pace. Bids arrive in unevenly spaced time intervals, sometimes coming fast and furious and other times just a trickle.

Then using the ongoing information and for the rate at which this information changes, such as the number of bidders, the frequency of the bidding, the changing in seller-buyer network to build a dynamic model.

The model operates during the live auction and forecasts price at a future time point during the auction as well as its final price.

|  |  |  |  |
| --- | --- | --- | --- |
| **Section** | **Description** | **Pages** | **Points** |
| **Introduction** | Why you are interested in the subject area that you selected for your project | 1 paragraph | 2 |
| **Problem Description** | What problem are you trying to solve? Who is the Customer? What will be the effect on the Customer? | 1 or 2 paragraphs | 3 |
| **Objective** | What is your objective | 1 or 2 sentences | 5 |
| **Data Curation** | Discuss your dataset(s): was data missing/ambiguous/poor quality? What were the challenges? How did you solve them. | 2 to 4 paragraphs | 10 |
| **Unsupervised Learning Results** | Look at some candidate datasets. Plot data, perform PCA, clustering, and other techniques. What did you find out about your candidate datasets? Why did you select the dataset(s) you did, and why did you select the outcome? | ½ to 1 page | 15 |
| **Supervised Analysis Results** | What techniques did you try and why? (Complete results in the appendix). Which worked the best? For the best technique(s), show output and graphs. | 1 page | 30 |
| **Conclusions** | Did your project meet the objectives? Why or why not? What conclusions can you draw from the results? | 1 to 2 paragraphs | 10 |
| **Follow-on Work** | How could you improve your project? What would you do differently? What could be a follow-on project? What did you learn of value? | 1 paragraph | 5 |
| **Appendices** | All your unsupervised and supervised analysis results. Show output for each technique you tried. | unlimited | 20 |