Online auction has been 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. Modelling and predicting the bidding price of online auction in real-time can be a challenging task. The price dynamics change throughout the auction and directly reflect the bidding behavior of all parties participated.

The original data set is from which was collected from Ebay. The dataset includes 10681 observations and 9 variables.

The dataset contains two different types of data, the bid history and the auction related information. The bid history is a sequence of bids placed over time, which can be considered as a time series type of data. One the other hand, the auction related information is a cross-sectional type of data. It contains product information, information about the bidders and the sellers and auction type information that does not change during the auction. Thus the dataset I am working on is 627 unique auctions with a sequence of their bidding histories.

Moreover, most of the time series methods typically assume that events arrive at equally spaced time intervals, for example in stock market, the timestamp of price is placed equal length. However, in the case for online auction data, the timestamp is generated by user activities, where events are not equally spaced. Therefore, each auction has different staring times, different ending times, different durations and different bidding decisions are recorded.

my intension here is to use smoothing splines and GAM to turn a discrete time series to a continuous function so that I can recreate equal length time series for each auction.

Also to notice that The real-time price we see on the website do not directly reflect the latest bid. The current price is the second highest bid+increment. People can bid as long as their bid is higher than current price. Therefore, as bid time increases, we sometimes see bids that are lower than previous bids.

To simulate the real-time auction price, I create a function following the logic of EBay auction displaying policy and transformed a sequence of bidding values to the sequence of 'current price'

For the learning part

The main challenge I encounter in this project is defining the machine learning object of my model. Although it's our intuition to predict the end-price of an auction, which can be done easily using given information such as its category, starting bid value and so on, the real challenge here how to include the ongoing information.

The main goal is to develop a model that can the auction price at a future time (T+1) using information from the present (T) and the past (T-1, T-2,…,1). Assuming the model parameters did not change from time T-1 to time T, at time T (present), the model can make a prediction about the future price at time T + 1.

For the baseline model I use the static features, that is the information are given before a certain auction starts and do not change during the auction. The static information includes:

- opening price of the current auctions
- duration of the current auctions
- sellers’ ratting of the current auctions

I trained the multiple linear model and KNN model

To capture the process of bidding I created a livebid with a 1 day lag as a feature for the prediction. The dynamic features also include the number of bidders along the auction, average bidding value, the standard deviation of the bidding value, the bid time difference and the bid value difference between each bid and the price velocity to capture the on-going features of an auction procedure. Adding this evolving information, we can create several dynamic models.

multiple linear model and KNN model XGBoost linear predictor

LSTM Neural Network

LSTM is a specific type of recurrent neural network which introduces a “memory cell” value that is passed down for multiple time steps and being sent in parallel with the activation output.

The dynamic models from previous section only use the information from the T-1.
With LSTM we can set a time window, a period of time to look back and predict the next target.
The neural network I built has three layers, the first one is a LSTM layer, the second is a dense layer with 32 knot and the third layer is a dropout layer to prevent overfitting and the fourth layer is a dense layer for the output.

Just like in any other neural network model, we have to rescale the input data to the range of the activation function. To compare the performance with other models, I revert the predicted values to the original scale and the MAE is 102.7578.

An ensemble of unsupervised clustering results and linear regression might have higher predictive capabilities, since the method of XGboost linear predictor also combines the tree classification models and the linear models.

My future works also includes the practical application of auction predicting models that helps the buyers making bidding decisions. Making bidding is further complicated since there existence of many auctions that offer the same or similar item simultaneously. Thus we might need to consider the competitive bidding behaviors as well.