30 October 2022

Article c - A Novel Method for Predicting the End-Price of eBay Auctions, 2013

The motivation

* Profit Maximization

How they do it

* Multinomial logistic regression (softmax), Naive Bayes (NB), and uniform prior Naive Bayes (UPNB) algorithms.
* For final price prediction, we find that for multiclass binary prediction decision tree models, UPNB is able to outperform both the general NB and softmax algorithms.

Interestingly

* The UPNB classifier uses item title as only input to predict if sale will occur.
* They decided to use records
  + I like this idea because in theory, two records should be identical

They suggest that

* Although price should reflect value, it does include variability due to factors such as Visibility, reputation, Start price
* Features stored included:
  + item number,
  + item title,
  + start time
  + end time,
  + text description,
  + detailed item specifications ,
  + seller feedback percentage and total score,
  + shipping costs,
  + return policy, and image presence.
* Each item was queried regularly through its complete auction duration so that page views, current price, and bid counts could be tracked as a function of time.
* They used Ebays API
* Calls were made daily

Before analysis

* Auction data is processed to remove potentially complicating outliers.
* Filter removes items where the complete text description (not just the title) contains any of the following words: skip, skips, damage, damaged, broken, warped, scuffs, and scratches. Previous work performed similar filtering to reduce their set by 25% [3]. This filtering reduces our data set by 30% and increases our UPNB prediction accuracy by 6-8% percent.
* Before performing this analysis, words that appeared only once amongst all titles were removed from the word list. Furthermore, words shorter than three characters, and the following common words, were removed from the list: and, for, from, that, there, these, this, and those. This analysis provided an error S = 0.25 when predicting on all genres.

Conclusion

* In reality we found that the start price is not always a good indicator of the end price. This can be attributed to the item’s intrinsic value or the flexible valuation of an item from user to user. These factors may be predominant in categories featuring collectibles.

My item

* They used records which in theory may be indistinguishable, however, surely a extra emphasis should be put on the sellers reputation as item may be easily damaged.
* How about Books?
* There are more than 1100 listing for just harry Potter and Philosopher's Stone?
* Or 1800 for Dan Brown’s Da Vinci Code.
* Or we can go the other way and look for hugely differentiated items, like bicycles. Huge number and huge range in quality.