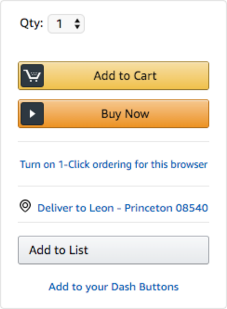
Predicting Sales Allocation to Third-Party Merchants

Leon Musolff[1]

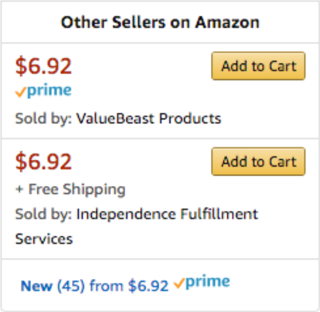
[1]Princeton University

April 2019

What is the Buybox?



(a) The Buybox



(b) The Featured Merchants Section

- One product on Amazon is typically sold by many 3P merchants
- The Buybox refers to 'Buy Now' button on an Amazon product page.
- Amazon allocates this to 3P merchants using proprietary algorithm

Why do we care?

- Online marketplaces have a lot of power in allocating sales.
  - can make or break a merchant
  - can influence prices directly
  - can influence pirces indirectly via quality of competition

- Online marketplaces may not have the right incentives.
  - Amazon gets paid fraction of revenue: want to keep prices high?
  - Amazon is not only marketplace, but also seller
  - Amazon gets more fees from merchants that use in-house fulfillment

The Data

- **Any Offer Changed** (AOC) notifications
  - AOC: a group of offers that belong to a product & condition
  - AOCs are sent whenever any of the offers changes
  - AOCs contain for each offer:
    - Buybox Status,
    - Featured Merchant Status,
    - FBA ("Fulfillment by Amazon") Status,
    - Listing Price,
    - Seller Feedback Count,
    - Seller Fraction of Positive Feedback,
    - Shipping Time.
- We have AOCs for 250 days; # per day varies (>2 million).
  - most results use 1 day (April 1st), interpretation uses 40 days
  - code to analyze all in parallel is written, but will take long to run

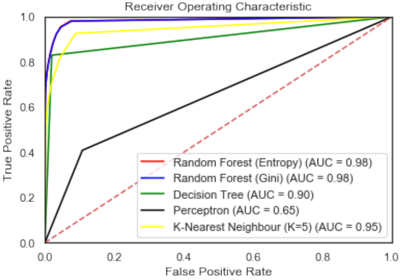
Preprocessing

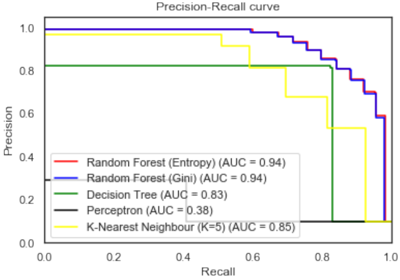
- We only keep Buybox eligible offers.
  - Algorithm determining eligibility is public.
  - We drop 16.92% of offers at this stage.
- Amazon may suppress the Buybox if no competitive offer.
  - To assess, would need prices of products on other market places.
  - As not available, drop these AOCs (10.73%)
- Amazon may assign two Buybox winners.
  - This happens if one of the winners is for 'Prime' customers.
  - By ignoring, we stack deck against finding an 'Amazon edge'.
  - So drop these AOCs (6.76%)

Feature Engineering

- What matters to whether an offer is in the Buybox are not covariates of the offer itself, but rather *how it compares to other offers* on these covariates.
- We thus generate comparison features, e.g. for all continuous vars:
  - difference to offer with lowest value of this variable?
  - difference to offer with highest value of this variable?
  - **difference to offer with lowest price on this variable?**
  - dummies: is this the lowest / highest offer on this variable?
- We also create comparison features for booleans, e.g. for FBA:
  - whether this offer and the k-th lowest offer are both FBA?
  - whether this offer is FBA and the k-th lowest offer is not?
  - and so on

Classifier Choice





- We briefly compare the performance of various classifiers.
- It is clear that the random forest outperforms other classifiers.
    - This is also the classifier used by prior work.
- Metrics not comparable to below as we use small sample for speed.

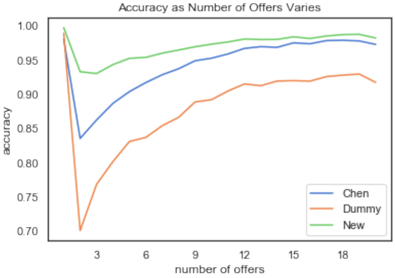
Prior Work: Chen et al I

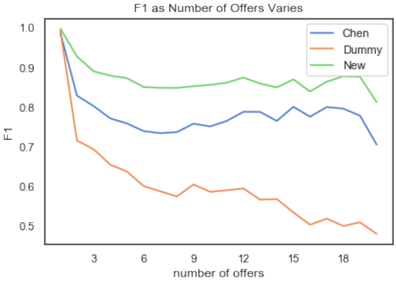
- The Buybox was previously studied by Chen et al (2011).
  - crawl 1000 best-selling products at 25min intervals for 3 months
  - data thus scraped; big issue: offers that do not vary
  - obvious sample selection issue (we also have one, but a different one)
  - do not observe shipping variables

Feature	**Importance** (Original Data)	**Importance** (Our Data)
Price Difference to Lowest	0.3711	0.2875
Price Ratio To Lowest	0.3402	0.2138
Positive Feedback	0.1031	0.0735
Feedback Count	0.0619	0.2547
Is the Product FBA?	0.0206	0.0651
Is Amazon the Seller?	0.1031	0.1054

Table: Feature Importances Using The Chen Classifier On Our Data.

Prior Work: Chen et al II

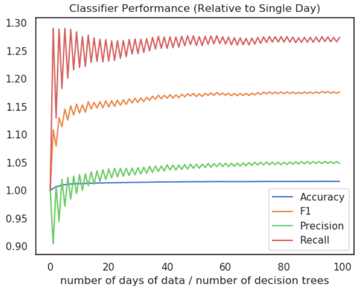


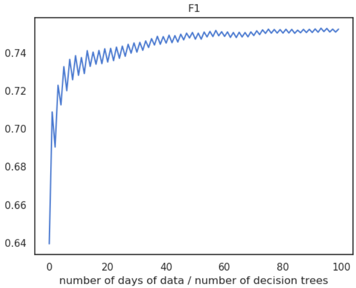


- We compare random forests employing Chen's feature set to our expanded feature set as well as a dummy classifier ("give Buybox to lowest offer")
- Both Chen and we outperform the dummy, but we outperform Chen at large margin due mostly to access to Shipping Time variables.

Big Data

- We now move on to using the entire dataset.
  - Logistical challenge: data does not fit into memory.
  - Solve by parallely training decision trees for each day on cluster.
  - Then merge decision trees to become random forest.
- The figures below show that classification performance increases as we use more days of data (but also more trees).





Feature Importances in Our Model

Group of Features	Weight
Landed Price	0.28
Listing Price	0.21
Shipping Time	0.13
Feedback Count	0.09
Shipping Availability	0.07
Positive Feedback (Fraction)	0.06
Positive Feedback (Count)	0.06
IsAmazon	0.03
Is Fulfilled By Amazon?	0.03
Shipping Origin	0.02
Number of Offers	0.01
Shipping Cost	0.00

Table: Relative Price is most important, but Shipping Time also matters.

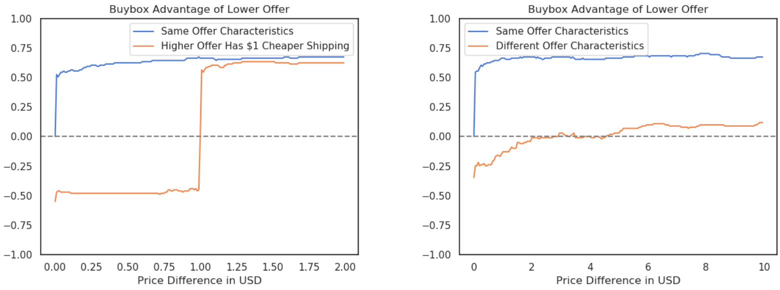
Interpretation: The Buybox Advantage

- We are interested whether Amazon only cares about price.
- Say there are exactly two offers, identical on all observables.
    - Presumably the cheaper offer is more likely to be in Buybox.
    - How does this advantage vary with observables?
- We define the Buybox advantage as

$$b := \mathbb{P}(\text{lower offer in Buybox}) - \mathbb{P}(\text{higher offer in Buybox})$$

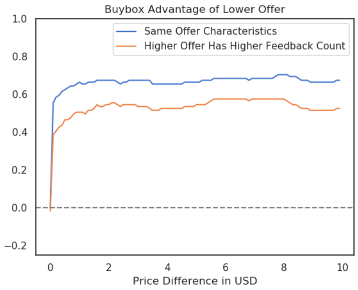
- Predictions
    - $b > 0$ for lower offer; b increasing in price difference
    - b decreases as we make the higher offer better in non-price covariates

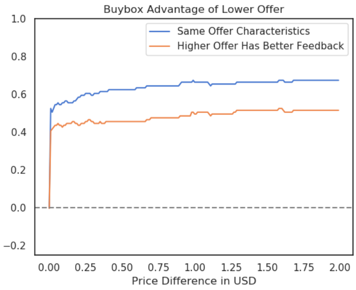
Interpretation: Shipping Speed Matters



- LHS: the landed price (not just the listing price) matters
- RHS: shipping speed matters a lot
  - in different offer characteristics case, higher offer ships faster

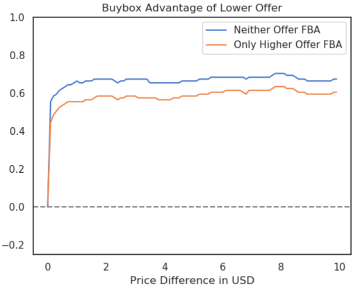
Interpretation: Seller Size & Quality Matter

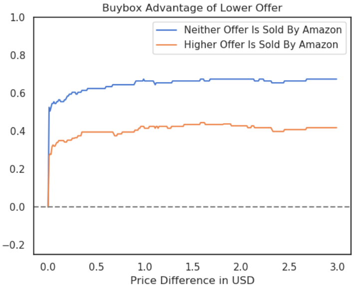




- On LHS, gradual phase-out of Buybox advantage for larger seller.
- On RHS, gradual phase-out of Buybox advantage for better seller.

Interpretation: The Amazon Edge





- Amazon is very slightly more likely to give FBA sellers the Buybox.
- Amazon is slightly more likely to give itself the Buybox.
- Caveat: Amazon has covariates we do not observe
    - plausible that Amazon / FBA sellers better on unobservables

Conclusion

- Amazon *typically* allocates Buybox to cheapest offer.
  - This is what you would do to keep prices low.
  - Hence, some evidence Amazon does not push prices up.
- But Amazon also takes into account other factors:
  - shipping speed seems extremely import,
  - and seller quality also matters.
- These other factors plausibly relate to customer experience.
  - Amazon is improving profit by giving customers what they want.
- Finally, Amazon seems to give itself and FBA sellers an edge.
  - Caveat: Amazon / FBA sellers may be better on unobservables.
  - This is worrying and warrants further investigation.

References

[1] Amazon. Annual Report: 2017. Technical report, April 2018.

[2] Le Chen, Alan Mislove, and Christo Wilson. An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. In *the 25th International Conference*, pages 1339–1349, New York, New York, USA, 2016. ACM Press.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[4] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1):25, Jan 2007.