Publications: ACM Recommendation system conference, Decision Support System (journal)

Amazon dataset: Cascades in reviews

1. Can’t see the purchase decision, only the reviews.
2. How first five reviews can predict the overall reviews
3. Windowed the reviews?
4. How about looking at examples of products with negative average rating (<2.5)?
5. Choose products with many reviews since: more reviews may be perceived as more objective and thus more trusthworthy.
6. Take into account not only the first few reviews, but also the time between reviews? The idea is that let’s say 2 items having the same number of reviews but for one item, the time of reviews are kind of spreading out evenly, while for another item, there might be a couple reviews and then a long interval in between there is no reviews at all, and then suddenly a bunch of reviews are present again.
7. Some kind of multiple linear regression model? The dependent variable is the average rating or the valance. The set of dependent variables are: average rating of the first 5 reviews, average rating of the next 5 reviews, etc. We can check out the estimated coefficients of the regression and see which batch of reviews tend to influence the average reviews the most.

SSCC datasets: Drugstore.com, stoneberry.com

1. A good thing about this data set: we can see the purchase decision too.
2. Is any of the previous results can be studied on other datasets too? Examine different categories too? Examine product of higher prices region?
3. Consider other independent factors:
   1. Review helpfulness
   2. Review recency (tie this to information cascades maybe?
   3. Has the customer purchased the product before? If yes, less likely to read review. Tie this with customer’s prior knowledge/attitudes, preferences about that brand.
   4. Non-linear effect of review characteristics. E.g., customer distrust on extremely positive reviews due to the existence of fake reviews.
   5. When customers did consult the reviews, which specific reviews and to what degree they read them?
   6. Include interactions of independent terms, thus need to use different models than GAMs.
   7. The content of the reviews: a sentiment analysis.
      1. 1-3 star reviews contain product shortcoming
      2. 3-4 star reviews contain suggestions for product improvement.
      3. Review’s readability, relevance, informativeness.
      4. A summary section under review tab: whether such summary is more influenced than individual reviews?
      5. Reviewer’s personal information, similarity between reviewers.
4. Can you explain a bit more on how the sentiment analysis of the product could be a possible thing we should look at?
   1. So last time we mention about
5. External information (prior information) of customer about the products’ quality/reviews’ trustworthiness.
6. Results:
   1. The inverted U-shape of purchased probability versus:
      1. Valance (reviews’ average rating): P[purchase] increases with valance starting from 3.0 but then show diminishing effect and decrease with valence about higher than 4.5
      2. Reviews’ length: longer reviews provide more information (more usage illustrations, more elaborations), but if it is too long, then P[purchase] decreases (due to more words create more chance of exposing counter-arguing in the review itself, or create overload of information).

I got the account and haven’t looked at the data yet, so I did not think about how we can proceed other than reading the papers you sent. Instead, I have worked on what I had worked with professor Berry so far.