THE UNIVERSITY OF CHICAGO

Thesis Proposal

How and Why Negative Reviews Are Useful? Assessing the Values of Negative Reviews on Amazon.com

By

Li Liu

December 2019

A proposal submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Dr. Philip Waggoner

Preceptor: Joshua Mausolf

Introduction

Product reviews, usually written by the customers who previously purchased the products, are important indicators for customer satisfaction and product qualities. You might spend some time reading product reviews when shopping online, but you may not be aware of how different reviews influence your final purchase decision. In particular, this paper will use the computational content analysis approach to show that the negative reviews (1 and 2 stars) are generally more informative and helpful than the positive reviews (4 and 5 stars) for products on Amazon because they cover more authentic information.

There are two major reasons for reading negative reviews. Firstly, they are more likely to reflect customers' true opinions than positive ones. Research has shown that people are more skeptical about whether the positive reviews are legitimate as companies inflate ratings by rewarding customers if they post a positive review (Dragan, 2016). By comparison, customers won't get any rewards if they write a negative review. As a result, negative reviews are less likely to be fake than positive ones (Beaton, 2018). Secondly, they tend to be more informative than positive ones. Customers might read the negative reviews carefully to predict what could go wrong and what's the likelihood of receiving a bad product (Beaton, 2018). Suppose you want to buy an iPhone 8, you already know Apple as a reliable brand and want to read reviews to figure out how others use the product's new features. Thus, positive reviews probably will not add much new information to your prior knowledge. Instead, negative reviews could show you some worse case scenarios that other customers have experienced, which could help you better decide whether it is worth purchasing.

This paper has many practical implications for individuals and businesses. As online reviews are prevalent in the digital world, understanding how and why negative reviews are helpful would enable them to use the reviews to make better purchase decisions. As for tech companies, such as Amazon, Ebay, Yelp, they could use this paper's insights to design personalized and informative interface for presenting the reviews to customers. As the amount of online review is increasing rapidly, this paper's methods could potentially help them extract and summarize the most relevant information from negative and positive reviews.

This paper uses a large data set of Amazon reviews and products to conduct the analysis. There are two parts of the analysis. In the first part, the focus is on how negative reviews are different in terms of the topics written by the customers. The main methods are topic modeling and exploratory data analysis. In the second part, the focus is on why negative reviews are different in terms of features that predict the helpfulness. I will find empirical evidence that certain features of the negative reviews are more helpful in customers' decisions. The main methods are text mining and supervised machine learning.

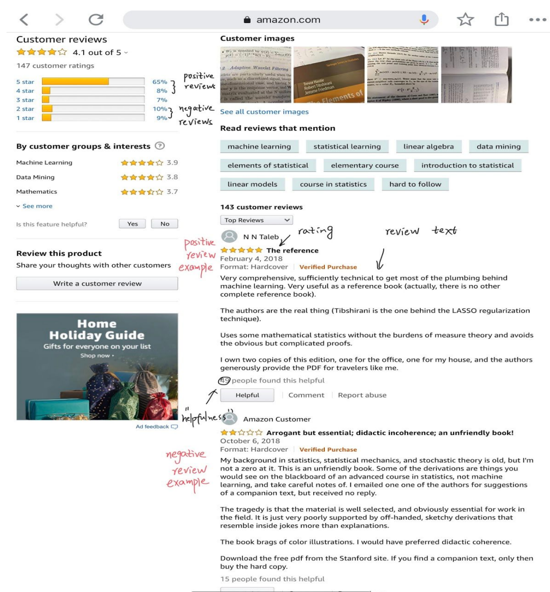
Literature Review

Broadly speaking, this paper contributes to the literature on product reviews and UGC (user-generated content) by discovering how negative reviews are distinctive to help customers' purchase decisions. Traditionally, in order to study what affects customers' purchase decisions, researchers would conduct surveys, focus groups, or lab experiments. Using content analysis methods, many recent papers in marketing and economics have illustrated that reviews tell a lot about customers. Timoshenko and Hauser (2019) show that user-generated content (including reviews, images, etc) could efficiently identify customers' needs by using machine learning. The text data has been popular in business and economics research as new features for the models (Gentzkow et al., 2019 and Berger et al., 2019). Thus, this paper serves as an empirical study that could help customers to make better decisions by informing them how to read product reviews wisely.

Data

The main data for this paper has two parts: the reviews and product metadata on Amazon. The dataset is scraped by several computer science researchers at the University of California, San Diego and made available for the public (Ni et al., 2019). There are two major advantages of using this data. Firstly, it has great breadth, as it contains 233 million reviews for various products in almost 30 product categories. Secondly, it has excellent depth, as it has reviews ranging from May 1996 to October 2018. As most of the existing literature only explore reviews in a short period for very few products, this data would enable me to conduct the comparative analysis on a large scale of products across times.

There are a lot of variables in the data. For each review, it has the reviewer's name, ratings, text, summary, attached image URL, and helpfulness votes (ratio of votes they received from all viewers), the label for verified purchase, etc. For each product, the data provides detailed information on its characteristics, such as color, price, package type, descriptions, technical details, similar products, image features, categories information, sales rank, etc. For this paper, I will start by analyzing the reviews for one popular product with a large volume of reviews, such as the book *The Elements of Statistical Learning*. After that, the same analysis could be easily applied to other books in machine learning and statistics.



Graph 1: Screenshot (with annotations) taken on 11/15/19 of Customers Review Page for the book: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition

Methods

There are two components of this paper. The first part uses topic modeling to answer how negative reviews are different in terms of the topics written by the customers. The second part applies machine learning to understand why negative reviews are different in terms of features that predict the helpfulness. To illustrate potential results, I will include hypothetical examples. The following paragraphs will describe the two parts in detail.

1. How are negative reviews more helpful?

In the first part, I will test the hypothesis that negative reviews cover a wider range of topics, such as product quality and customers' experience, compared with positive ones. I will start the analysis by using a subset of reviews data with just one product. To find whether negative reviews and positive ones contain different topics, I will use a type of content analysis called Topic Modeling. As a popular tool for analyzing text data, the topic model will group the words in the reviews into several latent topic groups. To find these latent topics, the model calculates the statistical correlations and groups the words that used together more frequently (Gentzkow et al., 2019).

In particular, as each review often contains several topics, the model should be able to assign different topic distributions to each review. Thus, I will use the Latent Dirichlet allocation, which is a specific type of probabilistic topic models (Blei and Ng, 2003). To select the number of topics, I will split the reviews randomly into training and testing sets with a ratio of 80:20. As a result, I can select the number of topics that enables the model, which is fitted with the training set, to predict the outcome best on the testing set. Finally, I will compare the overall topic distributions over the negative reviews and positive ones. For instance, the top three topics for negative reviews might be described as "unexpected problems", "missing components", "incorrect description", while the top three topics for positive reviews might be "work as described", "great value", "easy to use". Thus, negative reviews for this product show more informative critiques that might be contradictory to the seller's description and also not mentioned frequently in the position reviews. Although a customer won't necessarily encounter the same bad experience as described in the negative reviews, they would more likely to make better purchase decisions, especially when comparing different brands and deciding when to purchase.

2. Why negative reviews are more helpful?

The second part of the paper will build on the topics discovered in the first part. When reading a review, customers would evaluate the 'helpfulness' by reading the text of the reviews and vote by putting 'thumbs up' if they find a certain review is helpful. Thus, the helpfulness is measured by the ratio of people who vote for helpfulness against the total number of votes (Singh et al., 2017). To answer why negative reviews are helpful for customers differently compared with positive ones, I will identify the weights of features for predicting the helpfulness.

There are three sets of features that are relevant to this task. The first set of features are the binary variables for indicating whether the proportion of a topic in the review exceeds a certain threshold. The second set of features are linguistic or semantic characteristics in the reviews text that potentially affect customers' decisions process. They will be generated from the reviews data by text mining techniques. Some research have explored the significance of possible predictive features. Mudambi and Schuff (2010) used the regression model and found review extremity, review depth, and product type affect the review helpfulness. Textual features such as polarity, subjectivity, entropy, and reading ease are generated and proven to help predict the helpfulness by ensembles learning techniques (Singh et al., 2017). Besides from these features, I will add other features including sentiments, readability scores, grammatical mistakes, and so on. The third set of features are attributes, such as verified purchase labels, written time, images attached, reviewers' reputation, and so on.

For both positive and negative reviews, I will estimate the weights of these features in predicting the helpfulness by supervised machine learning. Mathematically, the customer's perceived helpfulness for reviews is the function of multiple feature variables (which are topics, textual characteristics, and attributes). However, as there is little known by the underlying structure of reviewers' perceived helpfulness, the standard multiple linear regression will not be able to make a good prediction. Thus, I will use penalized regression, which is a type of machine learning models, to find the feature importance. In penalized regression, the regression coefficients are shrunk towards zero by adding the penalized term to the loss function when estimating the coefficients (Hastie et al., 2009). In particular, Ridge regression adds an L2 term (sum of the squared coefficients) and Lasso regression adds an L1 norm (sum of the absolute coefficients). They will help to find the predictive features.

Here is an example of possible results from the second part of the paper. The top three weights for predicting the helpfulness in negative reviews might be verified purchase labels, the topic on "incorrect description", and length, while for positive reviews, the top three predictive features are the topic on "easy to use", images, and readability. This potential result illustrates that most predictive features in negative reviews are comparatively more likely to be authentic and sincere. As a result, negative reviews tend to be more trustworthy when we are critical about the product.

However, if you already have strong intention to purchase one certain product (such as buying an iPhone 8 as a gift for someone), you won't care what the negative reviews say and instead find the most readable positive reviews help you confirm that you are making an excellent purchase decision.

Reference

Ahn, D., Park, H., & Yoo, B. (2017). Which group do you want to travel with? A study of rating differences among groups in online travel reviews. *Electronic Commerce Research and Applications*, *25*, 105–114. https://doi.org/10.1016/j.elerap.2017.09.001

Beaton, C. (2018, June 13). Why You Can't Really Trust Negative Online Reviews. *The New York Times*. Retrieved from https://www.nytimes.com/2018/06/13/smarter-living/trust-negative-product-reviews.html

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2019). Uniting the Tribes: Using Text for Marketing Insight. *Journal of Marketing*, 0022242919873106. https://doi.org/10.1177/0022242919873106

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Dragon, L. (n.d.). Let's Talk About Amazon Reviews: How We Spot the Fakes. Retrieved November 18, 2019, from Wirecutter: Reviews for the Real World website: https://thewirecutter.com/blog/lets-talk-about-amazon-reviews/

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193

Gentzkow, M., Kelly, B. T., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *forthcoming*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon. com. *MIS Quarterly*, *34*(1), 185–200.

Ni, J., Li, J., & McAuley, J. (n.d.). *Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects*. 10.

Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research*, *50*, 67–83. https://doi.org/10.1016/j.annals.2014.10.007

Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., & Roy, P. K. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*, 346–355.

Timoshenko, A., & Hauser, J. R. (2019). Identifying Customer Needs from user-Generated Content. *Marketing Science*, *38*(1), 1–20. https://doi.org/10.1287/mksc.2018.1123