Reducing Search Cost of Online Reviews by Visualization

By Archana Date, Apurva Ranade, Vinamratha Veeresh, Sourabh Makker

**TO CITE THIS ARTICLE:**

Archana Date, Apurva Ranade, Vinamratha Veeresh, Sourabh Makker (2017), https://github.com/MakkerSourabh/

**FULL TERMS AND CONDITIONS OF USE:**

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [smakker1@student.gsu.edu](mailto:smakker1@student.gsu.edu)

The Publisher does not warrant or guarantee the article’s accuracy, completeness, merchantability, fitness for a purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

**ABSTRACTION**

Opinions, reviews, and consultation have long been recognized as a significant decision-making factor of the human behavior. When we do something that we are not entirely knowledgeable about, we tend to search for a reliable source of information in our network. The same human behavior applies when a consumer is out in the market to purchase a product. Traditionally consultations were collected from Word-of-mouth (WOM) sources, but with the development of the internet, WOM has moved from personal talks and small communities to vast consumer network engaged in feedback forms, online review systems and product rating systems where millions of users participate and share their opinions.

Our further investigation and research prospects is focused on *Reducing Search Cost of Online Reviews by Visualization*. We study how effectively these millions of reviews and consumer voices, can be presented in such a way that the search cost of the consumer is minimized.

For illustration purpose, we take an example of Amazon.com which is a lead player in online shopping and focusing on integrating customer reviews and opinions to enhance the product market. We will first see an overview of existing review system of Amazon.com - its components, its policies, its evolution over time and intended use.

**AMAZON COMPONENTS & POLICIES**

The Amazon review system is intended to provide customers with a holistic feedback of the product, based on experience of the customers who already own it, to facilitate decision making before shopping on Amazon.com. It consists of a mandatory star rating and an optional written review, and to review a product you need to be signed-in. You can modify or delete your review but can post only one review per product. Only when you are reviewing a product that you have purchased from Amazon.com, your review will display “Verified Purchase” against the review. For every product, an overall star rating is displayed which is the average of all the rating for the product out of 5. Also, the percentage of 1 start ratings through 5-star ratings are shown below the average rating. Amazon.com customer reviews are monitored for indecency, but do permit negative comments. For every review, there is an option for the reader to select if he/she found the review helpful or not and an option to “Report Abuse”. Based on this feedback of the reviews from the readers, the top positive and top critical reviews are displayed for every product. A customer can also leave comments for the posted reviews. There is a search bar to search through the reviews based on keywords or reviews for a product can be sorted based on “Top rated” or “Most recent” and filtered by “All reviewers” or “Verified purchasers” only, “Positive reviews”, “Critical reviews” or reviews based on star rating and by reviews containing text, video, images or video and images only.

**EVOLUTION OF AMAZON REVIEW SYSTEM**

In the initial days, any unregistered user could leave anonymous reviews of the products on Amazon.com. Starting from 2004, only registered users can leave reviews though they can stay anonymous by using a pseudo public name. Amazon allowed third party sellers and "marketers" to offer steep discounts or give away products in exchange for reviews. The basis for this program seemed innocent enough, get a product to try and leave an "unbiased" review but, as people wanted to continue to receive free or discounted products and stay in third party review programs, they left completely biased, overly positive reviews. This resulted in a product's page/listing rank based on the false positivity and the product rose to the top.

In 2015, Amazon rolled out a new review platform based on machine learning. Developed in-house, this system gave “more weight to newer reviews, reviews from verified Amazon purchasers, and those that more customers vote up as being helpful.” In 2016, Amazon prohibited incentivized reviews unless they are facilitated through the Amazon Vine program. Amazon – not the vendor or seller – identifies and invites trusted and helpful reviewers on Amazon to post opinions about new and pre-release products; there is no incentive for positive star ratings, no attempt to influence the content of reviews, or even require a review to be written; and there is a limit the total number of Vine reviews that are displayed for each product. Vine has important controls in place and has proven to be especially valuable for getting early reviews on new products that have not yet been able to generate enough sales to have significant numbers of organic reviews.

**LITERATURE REVIEW**

**Highlighting the Importance of Online Review System:**

There are various studies conducted in recent years which demonstrated that online product ratings and consumer reviews have great impact on the consumers making purchase decisions. More and more potential buyers are increasingly relying on the online information in the form of feedback form and review system for product and service information. *(Chevalier and Mayzlin 2006; Clemons,Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007). Moe, W. W., and Trusov, M. 2011. "The Value of Social Dynamics in Online Product Ratings Forums," Journal of Marketing Research (48:3), pp. 444-456-456.*

Reduce Consumer Search Cost:

There are different challenges that the customer faces in the existing review system. The one challenge that has come across having most important impact is the Search Cost. Search cost means searching the relevant information from the reviews. *Chen, Pei-Yu; Wu, Shin-yi; and Yoon, Jungsun, "The Impact of Online Recommendations and Consumer Feedback on Sales" (2004). ICIS 2004 Proceedings. 58.*[*http://aisel.aisnet.org/icis2004/58*](http://aisel.aisnet.org/icis2004/58)*.* Reduction in consumer search costs has important implications on firm strategy and consumer behaviors.

Decomposing Reviews:

Reviews are written by non-technical consumers who may not provide the required analysis pattern while evaluating or reviewing a product. Such reviews may be biased due to self-selection of the reviewer and may fail to provide the complete and accurate information about the quality. Therefore, it becomes important to decompose the individual reviews and the rating system to quantitative components which can be mathematically evaluated.

[*https://doi.org/10.1287/isre.1070.0154*](https://doi.org/10.1287/isre.1070.0154)*, Xinxin Li, Lorin M. Hitt, (2008) Self-Selection and Information Role of Online Product Reviews. Information Systems Research 19(4): 456-474*

Figure Out the Negatives and Positives of the Review:

In most of the competitive products, every product has a good feature and a bad feature. Reviewer may get affected more by a feature which he priorities more than the other and his reviews may be biased because of his personal preferences. Such reviews may be useful for a smaller group of people with similar taste but may give a wrong notion about the over-all product quality for the much larger mass consumer group. For example, A consumer might provide negative feedback and less rating to a product which was received damaged but it is shipping and delivery issue and not the product issue. *Sidharth Muralidharan, Hye Jin Yoon, Yongjun Sung, Jessica Miller & Arturo Lee (2017) Following the breadcrumbs: An analysis of online product review by online, Journal of Marketing Communications, 23:2, 113-134, DOI: 10.1080/13527266.2014.949824*

Reputation and Opinion Extraction:

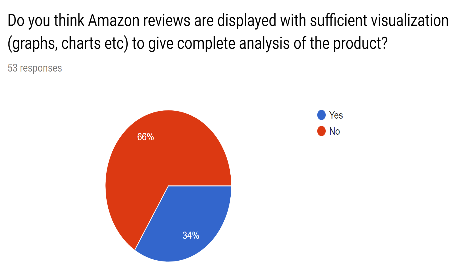
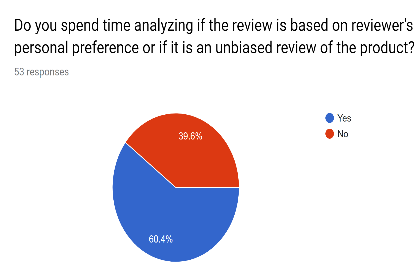
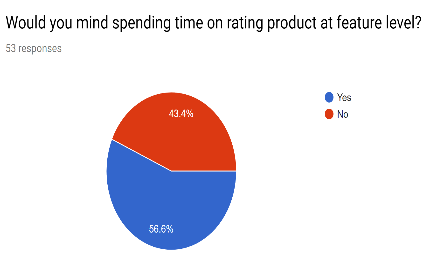
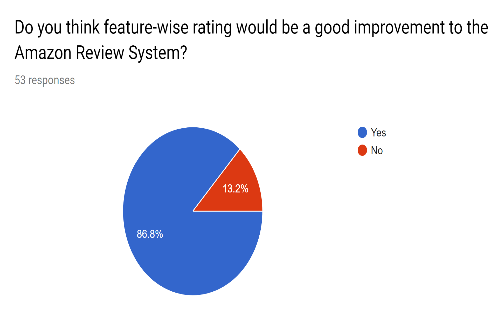
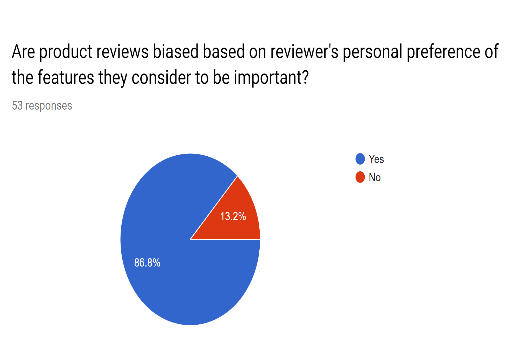
To quantitatively summarize the quality of a product, it becomes important to visually improve the data for the customer. Data mining and Visualization has already been proved strategically important in the field of online shopping and sentiment analysis. It consists of opinion extraction with fundamental characteristics like word extraction, feature highlighting, typical word extraction, and corresponding analysis. *Mining Product Reputation on the Web,* [*https://dl.acm.org/citation.cfm?id=775098*](https://dl.acm.org/citation.cfm?id=775098)*,* [*Satoshi Morinaga*](https://dl.acm.org/author_page.cfm?id=81100394130&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685)*,* [*Kenji Yamanishi*](https://dl.acm.org/author_page.cfm?id=81100329516&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685)*,*  [*Kenji Tateishi*](https://dl.acm.org/author_page.cfm?id=81100281613&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685)*,* [*Toshikazu Fukushima*](https://dl.acm.org/author_page.cfm?id=81100319620&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685)

**EXISTING PROBLEM IN AMAZON REVIEW SYSTEM**

The increased usage of social media platform to share, recommend, rate, and purchase products provided by Online shopping websites like Amazon.com, Flipkart.com, Shein.com etc. has increased the available information multifold. In this scenario, the problem is too much data and it is presented on web in a way which is not supporting the customer decision to the extent which it could possibly do.

Our study provides affirmative answers to the most important question here;

* Are online reviews visually informative?

We conducted a survey where graduate students of Georgia State University’s Computer Information Systems department were asked if they could make a decision on product quality, with little search cost, looking at the ratings and review system of Amazon.com. 

The two major reasons why existing Amazon.com review system may fail to provide complete information about the quality of the product are;

1. **High Search Cost:**

The current review system is a mix of positive reviews, negative reviews and neutral reviews. There is a lot of search cost involved for the consumer to read tons of reviews and understand what the reviewer must say. In this fast pace life, consumer rarely spends so much researching on the product and ends up buying what looks like the best fit for him. This leads to wrong purchases and customer dissatisfaction.

1. **Feature Level Rating/Visual impact Not Available:**

In the current scenario, a reviewer rates the product at an overall level but his rating could be based on selected features and not all the features of the product. This self-selection of the properties on which early buyers create reviews and rate the product can create bias in opinion. <https://doi.org/10.1287/isre.1070.0154>*, Xinxin Li, Lorin M. Hitt, (2008) Self-Selection and Information Role of Online Product Reviews. Information Systems Research 19(4): 456-474*.

DATA AND METHODOLOGY DESCRIPTION

Existing Visual representation of review on Amazon.com:



To get an idea of a product, if a customer decides to read all 5 stars and 1-star reviews it will be 81% reviews which means approximately 743 reviews. This is a time-consuming process. Let’s consider another scenario, a customer is only interested in what is the sound volume level and quality of this musical book but not all review contains that info so he must search through all reviews to check the review of these sound features.

**Feature Level Rating:**

We propose capturing feature level review from the user. This would be a structured data approach. The review section would be modified and the user would be provided with a new template where the product would be required to be rated at feature level. The list of the product features would be predefined as per the category of the product.

The more evolved rating at feature level would look the detailed chart below;

|  |  |  |
| --- | --- | --- |
| Feature List | Rating (out of 5) | Recommendation |
| Overall Rating (system calculated) | 3.3 | 67% |
| Sound Quality | 4 | 80% |
| Volume | 4 | 80% |
| Design | 1 | 20% |
| Content | 3 | 60% |
| Safety | 4 | 80% |
| Product Delivery | 4 | 80% |
| Write review here |  | |
|  |  |  |

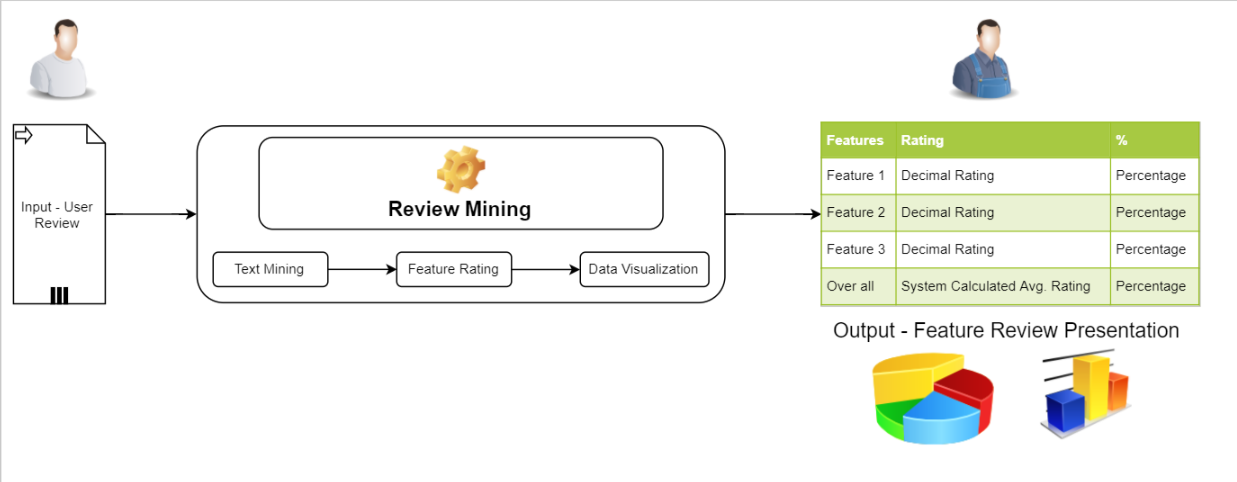
Description-

Overall average product. Volume feature is good, it is safe to use, content is average and design is bad.

After calculating the average rating for each feature, summary of all feature level reviews can be shown visually using a graph.

Reviews Mining:

As suggested in the structured data approach, where the review input form would be modified and user would be provided with a new template to rate the products at feature level, as an improvement, Data Mining techniques could be applied to self-analyze the sentiment of the review and rate the features at real time. This system rating would help reviewer save rating time but also provide a flexibility to overwrite the system generated rating for his review.



Data Collection:

For mining existing reviews, review data would be collected from credible datasets available. The data would be preferably collected in CSV/JSON format for easy readability and compatibility with Mining Tools. Few available datasets are <http://jmcauley.ucsd.edu/data/amazon>;

https://www.kaggle.com/c/rsm-bda-predict-online-review-scores

Data Preprocessing:

The data from the dataset would be preprocessed for filling missing values, reducing noisy data, and removing inconsistency. Data harmonization would be performed to elaborate short codes like (st, rd, etc) to actual words like (street, road, etcetera)

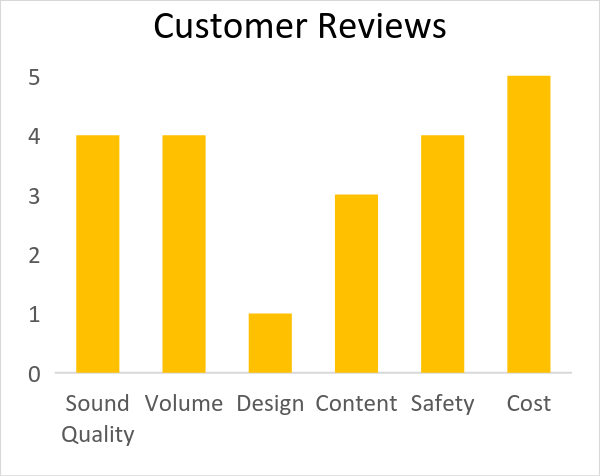
Mining Clean Data:

The cleaned data would be evaluated using Data frames, python libraries, pandas and the frequency of features would be calculated and analysis would be performed to see if the given feature is referred in positive sense or negative sense. This analysis would help in rating the feature of the product.

User Interface:

The analyzed data, which is now the ratings of the product at feature level would be presented on the website using matplotlib, Tableau and R to provide Visual summary of the reviews.

User Interface of Feature-level Review Summary:

By looking at the graph above, customer will make decision whether he should buy a product or not if he wants good design then he will not buy it but if he is interested in content and sound quality then he will surely buy it.

**Visually capturing evolving Sentiments for the Product:**

When any product is launched many experts and test customers review it and those opinions may not be representative of the opinions of the broader consumer population in later time periods. With time opinion, options and choice of people changes hence reviews in previous years may not be valid in this year. For example, a consumer may have found some product useful 2 years ago but at present he has better options available. It is better that a customer should get recent reviews first to eliminate effects of older reviews, and also a comparison between ratings of early bias reviews and overall average rating of the reviews. A graph showing average review rating of product per year starting from product launched year can be shown.



**REFERENCES**

Olbrich, R., & Holsing, C. (2011). Modeling consumer purchasing behavior in social shopping communities with clickstream data. International Journal of Electronic Commerce, 16(2), 15–40.

Chevalier, J., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. J. Marketing Res. 43(3) 345–354.

Godes, D., D. Mayzlin. 2004. Using online conversations to measure word-of-mouth communication. Marketing Sci. 23(4) 545–560.

Xinxin Li, Lorin M. Hitt, (2008) Self-Selection and Information Role of Online Product Reviews. Information Systems Research 19(4):456-474.

Mining Product Reputation on the Web, [Satoshi Morinaga](https://dl.acm.org/author_page.cfm?id=81100394130&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685), [Kenji Yamanishi](https://dl.acm.org/author_page.cfm?id=81100329516&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685),  [Kenji Tateishi](https://dl.acm.org/author_page.cfm?id=81100281613&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685), [Toshikazu Fukushima](https://dl.acm.org/author_page.cfm?id=81100319620&coll=DL&dl=ACM&trk=0&cfid=819729828&cftoken=95860685)

Sidharth Muralidharan, Hye Jin Yoon, Yongjun Sung, Jessica Miller & Arturo Lee (2017) Following the breadcrumbs: An analysis of online product review by online, Journal of Marketing Communications, 23:2, 113-134, DOI: 10.1080/13527266.2014.949824

Chen, Pei-Yu; Wu, Shin-yi; and Yoon, Jungsun, "The Impact of Online Recommendations and Consumer Feedback on Sales" (2004). ICIS 2004 Proceedings. 58.<http://aisel.aisnet.org/icis2004/58>*.*