Estimation and Analysis of Usability Scores of Product Reviews

Aditee Vasant Jadhav, Aditi Khurd, Krupa Dhirajlal Vadher and Pranav Sudhir Dixit

Abstract— **The recent advancements in technology have enabled every industry to showcase their products on their websites and due to such large amount of choices, online shopping is turning out to be a popular choice of people, as they can do it easily at home without being have to move out physically. People look at product reviews and then make decisions of their choice. There are huge number of reviews on e-commerce websites even for less popular products. The consumers and businessmen find it a big challenge to process such a big data of reviews. Thus customers find it difficult to go through all reviews and many good reviews get buried down huge number of reviews posted afterwards. To solve the problem of ignoring good reviews, we propose to evaluate the product reviews and calculate the usability scores of the reviews using different NLP techniques. We define usability scores of the reviews as the average normalized score calculated using polarity, subjectivity, readability scores of the review. This project will help fair chance of getting viewed to the good quality reviews. We suggest that assigning usability scores to the reviews will help reviewers to write better reviews and help customers make better buying decisions by looking at better quality reviews.**

I. INTRODUCTION

Online shopping has gained popularity over decades due to ease for customers to buy the products sitting comfortably at home and relying heavily upon other user reviews for the products. The product reviews play a vital role in helping a customer to make buying decisions. Thus, researchers have been attracted towards understanding the role of product reviews in businesses. These product reviews are useful to both decision making of customers and quality improvement for industrial firms. The advancements in technology and increased users of internet has contributed towards a rapid increase in online product reviews being posted and being looked at.

The product reviews can make both positive and negative impact on the customers. This impact of reviews helps in better decision making for the customer. However, the large amount of product reviews make it infeasible for the customer to go through all the reviews and then decide on product purchase. Also, not all reviews contribute a helpful information in decision making process of the customers and if such reviews are displayed on the top pages, the customer may go in dilemma towards product buying decision. Thus, it would be great if a customer could get an insight of helpful reviews that represent an overview about product and its features. Some e-commerce companies like Amazon.com provide an option of voting helpful reviews to the customers. The reviews on Amazon.com are by default sorted by using the helpfulness voting received. Thus, customers can make better buying decision and this helps increasing the product sales. However, there are certain problems in sorting reviews based on the helpfulness voting received. One of the problems is that a very small fraction of product reviews receive helpfulness voting. Another problem is the situation where recently posted reviews do not get votes. One more problem is faced when the reviews are posted at such a speed that some useful reviews may get buried down the pile of huge amount of reviews before receiving any helpfulness voting. It is also worth noticing that the reviews with helpfulness voting are exposed highly to the users and they gain even more visibility to the users. This resembles the real world phenomenon of 'rich getting richer and poor getting poorer'. Another scenario that needs attention is that the customers looking at the reviews with high helpfulness voting are influenced without noticing the review posting date and the context of the review.

Although the product reviews are useful to the customers in the process of purchase decision making, extracting the helpful reviews from a huge dataset is a tedious task. The reviews data is increasing in size day by day at a great speed. Thus, researchers suggested applying machine learning techniques for extracting useful information out of data. Many sentiment mining algorithms and natural language processing techniques have been developed to extract helpful information out of given reviews. Still the huge size of reviews needs to be processed every time we dig into reviews dataset. Also the existing experiments have focused on sentiment mining and not much work is done on measuring the quality of the review. In our project, we focus on measuring the quality of reviews and thereby estimating the helpfulness of the reviews instead of relying solely on helpfulness voting. We propose to calculate the quality of the review in terms of usability score which will be computed based on different techniques of assessing the textual reviews like polarity, subjectivity, readability index, etc. We will make use of these metrics and aggregate them to discover the helpfulness of the reviews.

This report is organized as: Previous research done in this field and how our work differs from others is explained in section II. The details of the methods we will apply and expected results are explained in section III.

II. RELATED WORK

Numerous research work is done for determining the helpfulness of the product reviews. Liu et.al.[1] proposed a system that treats helpfulness of reviews as a binary classification issue. Their system labels reviews as 'favourable' and 'unfavorable' based on vectors of review text. However, they did not use helpfulness voting of the reviews in their system. Ghose et.al. [2] proposed a system that assigns two ranks to the reviews based on either helpfulness voting received or based on effective sales of the product. They analyzed the review data along with its effect on sales. They discovered that the products that receive the reviews that have subjective or objective text tend to have more sales.

Otterbacher et.al. [3] suggested machine learning model for predicting the helpfulness voting of the reviews based on factors like review length, style of writing of the user. The authors claim that use of different metrics for determining quality of review will help in efficient and more accurate determination of helpful reviews. Danescu et.al. [4] proposed a model to establish relationship between ratings of the products and fraction of reviews that receive helpfulness voting. The authors concluded that the helpfulness voting is generally associated with average product rating.

Mudambi et.al. [5] conducted experiments to determine what properties of review text make it more prone to receive helpfulness voting. The authors found out that moderate reviews received more helpfulness votes than extreme reviews. They also discovered that length of review and subjectivity impacted helpfulness of the review in a greater way. Huang et.al. [6] carried out experiments to establish the equation for impact of review length on review helpfulness. The authors found that length of review measured in term of count of words had limited impact on helpfulness of the review. When number of reviews in the system cross certain threshold, the length of review shows lesser impact on helpfulness. The authors suggested that extracting semantic features from reviews might discover new patterns that may contribute towards estimating helpfulness of the reviews.

The literature survey here suggests that not all the reviews on a product are seen by customers and top reviews receive more helpfulness votes and low reviews get comparatively less votes. This elevates the Matthew social effect of "rich getting richer and poor getting poorer". Thus is has become necessity to understand the semantic features of reviews which may contribute towards helpfulness of the reviews. Thus, it would be great if we could compute helpfulness of reviews using semantic features rather than waiting for customers to cast their votes and thereby rank reviews according to calculated helpfulness of reviews.

III. METHODOLOGY

The facts and the issues stated in literature survey makes an observation that the Matthews social effect hampers the estimation of helpfulness of reviews. In this project, we plan to address this issue by calculating the helpfulness of reviews in terms of 'usability score'. The term 'usability score' cab be defined as an aggregated value of different semantic features of the reviews. We will first compute different semantic metrics for reviews and final output will be aggregated value of all semantic metrics. The methodology we plan to apply and the metrics we plan to use in our project for computing semantic features of reviews can be listed and described as follows:

A. Dataset

For this project we plan to use Amazon reviews data which readily available over internet. Two of our teammates had worked on recommendation systems competition for kaggle and they have worked on this dataset. However, they had used star ratings for their experiments and they did not work on text reviews for recommendation of products. They suggested that taking a deep dive into this reviews dataset may help us discover new patterns and hence we decided to work on Amazon reviews dataset. Their previous work focused on product id, user id and start ratings in terms of numeric value. When they worked on this competition, they thought that although they did not work on text reviews, it might have hidden knowledge. When we decided to work on this dataset, we observed that this dataset is very large and just finding vector of words in a review will not be useful for recommendation system. Literature survey also showed that quality of reviews is an important factor. Hence we decided to work in finding helpfulness of reviews from different semantic features.

Amazon reviews dataset is very large and it would be time consuming to work on reviews of each category of products in this data. Hence, we decided to work on a subsection of Amazon reviews dataset: reviews dataset on toys and games. Toys reviews dataset contains field like ASIN (Amazon Standard Identification Number), reviewer id, reviewer name, review time, review text, helpfulness votes, ratings. Out of these attributes, we plan to first focus on computing semantic features from reviews text field only. Then we will try to analyze relation between semantic features and ratings.

B. Semantic features

In this project we focus on first computing different semantic features of text reviews and then we will aggregate them together into a new term defined as usability score'. The semantic features in this project are: subjectivity, polarity, flesch index, entropy, Dale Chall index, lex diversity, helpfulness ratio, review length and set length.

The 'usability score' of review is calculated by taking aggregation of all semantic features stated above. The 'polarity' of review is calculated by subtracting negative score of review from positive score of review. These positive and negative scores can be calculated using existing databases like SentiWordNet. The 'polarity' score of the review will tell if the review states positive information or negative information.

The 'subjectivity' score of the review is defined as the fraction of review statements that express an opinion against total number of statements in the review text. This score is calculated using number of nouns, adjectives, verbs in review text. The 'review length' and 'set length' is calculated by measuring the number of total words in the review and the number of unique words in the review respectively. The 'lex diversity' is computed by dividing 'set length' by 'review length'.

The 'readability' score of the review tells about level of ease of language used in the review. To calculate readability score, we use two different metrics: Flesch index, Dale Chall index. The 'entropy' of the review is the probabilistic model representing amount of information embedded in the review.

C. Data preprocessing

For our project, we are using only text reviews from Toys reviews dataset which is a subset of Amazon reviews dataset. The text reviews are created by customers and are heavily dominated by their local language. There may be lot of noise in text reviews. To reduce this noise, we plan to perform certain preprocessing steps on the data as explained follows.

We will first remove missing values by eliminating 'Nan' from dataset. Then we will convert all the words in text review to lowercase. We plan to remove the words that do not contain any letter from the review. We will also remove the words containing numbers from the review text. The punctuation symbols from text review will also be removed. Then we will discard the reviews whose length is below certain threshold level. Because less descriptive reviews do not contribute as helpful reviews.

IV. CONTRIBUTIONS

The planned individual workload of our project teammates towards project is as:

- Aditee Jadhav: Code for - Calculating 'subjectivity score', calculating 'flesch index'.
- Aditi Khurd: Code for - Calculating 'lex diversity score'
- Krupa Vadher: Code for - Calculating 'polarity score', calculating 'entropy score'.
- Pranav Dixit: Code for - Calculating 'Dale Chall index', calculating 'helpfulness ratio'.

The individual contribution towards project completed till now is as:

- Aditee Jadhav: Gathering dataset, Project report writing, Code for - Data conversion, Data visualization, Calculating 'subjectivity score', calculating 'flesch index'.
- Krupa Vadher: Code for - Data preprocessing, Calculating 'polarity score', calculating 'entropy score'.

The rest tasks of individuals will be accomplished till next milestone.

V. CONCLUSIONS

Product reviews play a vital role in buying decisions of customers and thus it helps firms to grow their business. However, not all reviews contribute towards this goal and hence we need to establish a meaningful way of computing helpfulness of the reviews, so that customers can see more helpful reviews and make better buying decisions. This will also help users in writing the better reviews. In our project, we compute helpfulness of reviews in term of 'usability score' which is a aggregated representation different semantic features of review text. To define semantic features of review text, we compute subjectivity score, polarity score, flesch index, entropy score, dale chall index, lex diversity, review length and set length. We expect to establish correlation between helpfulness of the reviews and semantic features.

ACKNOWLEDGMENT

The authors are grateful to Prof. Rojas Carlos for providing a valuable feedback to our project and his comments have helped us write better project report.

REFERENCES

[1] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, M. Zhou, "Low-Quality Product Review Detection in Opinion Summarization", Proc. Joint Conf. Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 334-342, 2007.

[2] A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," in IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 10, pp. 1498-1512, Oct. 2011.

[3] J. Otterbacher, "Helpfulness in Online Communities: A Measure of Message Quality", CHI '09: Proc. 27th Int'l Conf. Human Factors in Computing Systems, pp. 955-964, 2009.

[4] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, L. Lee, "How Opinions Are Received by Online Communities: A Case Study on Amazon.com Helpfulness Votes", Proc. 18th Int'l Conf. World Wide Web (WWW '09), pp. 141-150, 2009.

[5] Mudambi, S.M., Schuff, D., "What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com", (SSRN Scholarly Paper No. ID 2175066), Social Science Research Network, Rochester, NY, 2010

[6] Huang, A. H., Chen, K., Yen, D. C., & Tran, T. P., "A study of factors that contribute to online review helpfulness", Computers in Human Behavior, 48, 17–27, 2015.