Characterizing and Predicting Early Reviewers

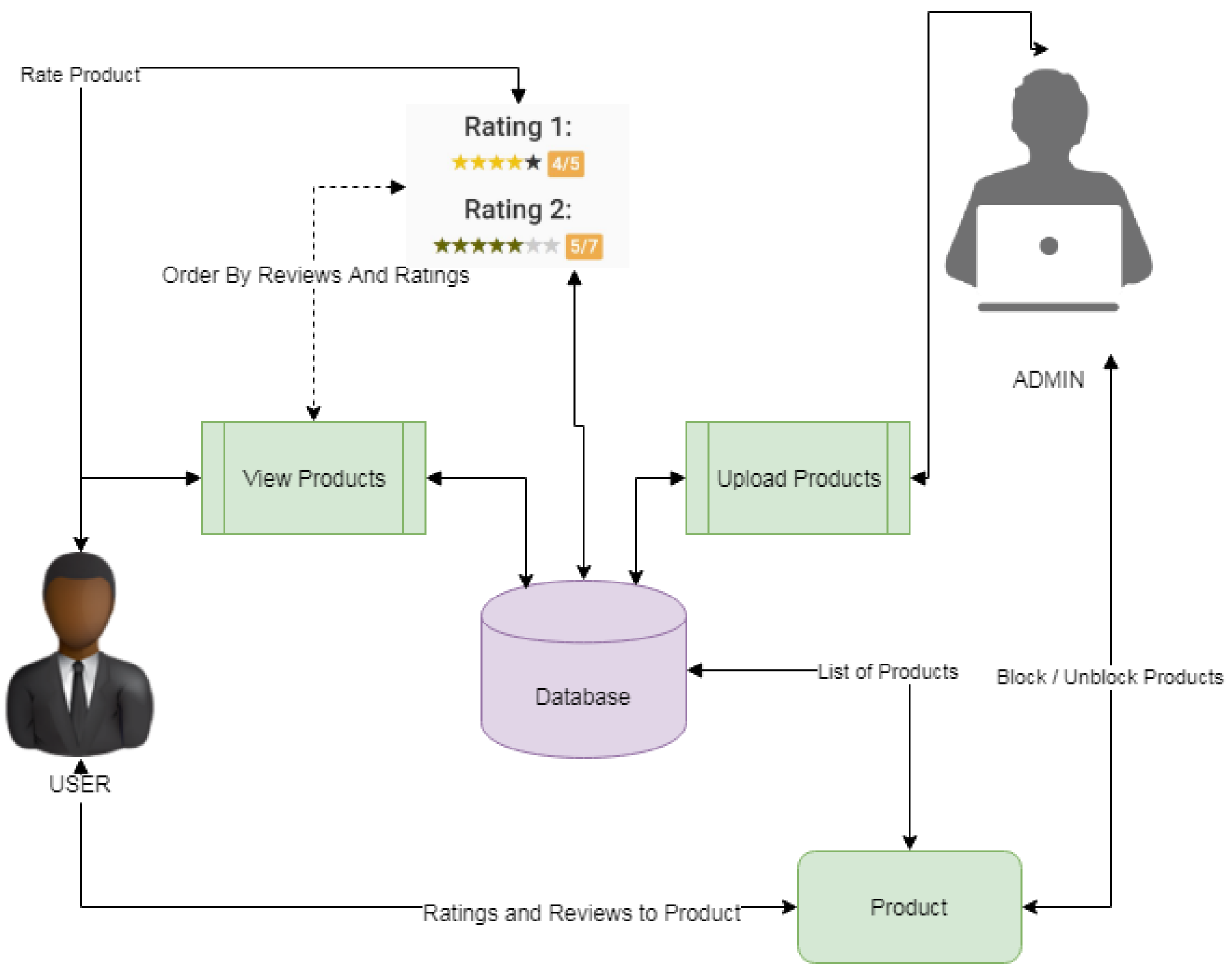
Sk. Heena #1, Beulah #2,Syama Sindhuja #3, Maheswari #4,Meghana

#1Asst. Professor, Dept of Computer Science and Engineering, Qis College of Engineering and Technology, Ongole
#2Student, Dept of Computer Science and Engineering, Qis College of Engineering and Technology, Ongole
#3 Student, Dept of Computer Science and Engineering, Qis College of Engineering and Technology, Ongole
#4 Student, Dept of Computer Science and Engineering, Qis College of Engineering and Technology, Ongole
#5 Student, Dept of Computer Science and Engineering, Qis College of Engineering and Technology, Ongole

ABSTRACT:
Online reviews have become an important source of information for users before making an informed purchase decision. Early reviews of a product tend to have a high impact on the subsequent product sales. In this paper, we take the initiative to study the behavior characteristics of early reviewers through their posted reviews on two real-world large e-commerce platforms, i.e., Amazon and Yelp. In specific, we divide product lifetime into three consecutive stages, namely early, majority and laggards. A user who has posted a review in the early stage is considered as an early reviewer. We quantitatively characterize early reviewers based on their rating behaviors, the helpfulness scores received from others and the correlation of their reviews with product popularity. We have found that (1) an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our analysis of product reviews also indicates that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity. By viewing review posting process as a multiplayer competition game, we propose a novel margin-based embedding model for early reviewer prediction. Extensive experiments on two different e-commerce datasets have shown that our proposed approach outperforms a number of competitive baselines.

1 Introduction

The emergence of e-commerce websites has enabled users to publish or share purchase experiences by posting product reviews, which usually contain useful opinions, comments and feedback towards a product. As such, a majority of customers will read online reviews before making an informed purchase decision . It has been reported about 71% of global online shoppers read online reviews before purchasing a product . Product reviews, especially the early reviews (i.e., the reviews posted in the early stage of a product), have a high impact on subsequent product sales . We call the users who posted the early reviews early reviewers. Although early reviewers contribute only a small proportion of reviews, their opinions can determine the success or failure of new products and services . It is important for companies to identify early reviewers since their feedbacks can help companies to adjust marketing strategies and improve product designs, which can eventually lead to the success of their new products. For this reason, early reviewers become the emphasis to monitor and attract at the early promotion stage of a company. The pivotal role of early reviews has attracted extensive attention from marketing practitioners to induce consumer purchase intentions . For example, Amazon, one of the largest e-commerce company in the world, has advocated the Early Reviewer Program1, which helps to acquire early reviews on products that have few or no reviews. With this program, Amazon shoppers can learn more about products and make smarter buying decisions. As another related program, Amazon Vine2 invites the most trusted reviewers on Amazon to post opinions about new and prerelease items to help their fellow customers make informed purchase decisions.



2 Related Work

Products with Complete Lifetime

Product Review Time Span refers to the time span between the first and last received reviews for a product. Formally, given a product p, its product review time span is the range between the timestamps of its first and last reviews, i.e., [s1, sNp]. Our observation window is defined as the period between the start and end time of datasets.

Amazon dataset contains product reviews ranging from May 1996 to July 2014, and Yelp dataset contains product reviews ranging from July 2004 to January 2017. The observation windows are 18 years and 13 years respectively. It might be the case that some products have their reviews falling outside of our observation window. We propose the following strategy to determine whether a product's review time span is complete within our observation window.

Determining the Complete Review Time Span

We first introduce the concepts of leading gap and trailing gap. Given an observation window [sstart, send], the leading gap of a product p, denoted by Δ (p) L , is defined as the time difference between s1 (when the first review was found within the observation window) and sstart, while the trailing gap Δ (p) T of a product p is defined as the time difference between sNp (when the last review was found within the observation window) and send. Our key idea is that if the maximum interval between two consecutive reviews of a product p is smaller than both the leading and trailing gaps of product p, then we have observed a complete review time span (e.g., Figure 1(a)). However, if it is not the case, then we have only observed a partial product review time span (e.g., Figure 1(b)). We denote the maximum interval for product p by Δ (p) M , and it is computed as: Δ (p) M = max Np−1 i=1 si+1 − si . Based on our idea, we consider the lifetime for product p is complete if it satisfies: Δ (p) L > Δ (p) M and Δ (p) T > Δ (p) M . 2.1.2 Estimating the Product Lifetime Given a product, we take its complete review time span as a proxy measure of its lifetime. It should be noted the time Fig. . An illustrative example on the complete and incomplete review time span for a product. In our observation window, product p1 has a complete review time span while product p2 has an incomplete review time span. Green triangles indicate the observed boundaries of product review time span, and red triangles represent reviews which are outside the observation window. span derived from product reviews may not exactly align with the actual product lifetime from a customer's point of view, i.e., the period of time over which a product is first brought to market and eventually removed from market. Since our current datasets do not contain any explicit purchase information, it is not possible to accurately derive the product lifetime. Nevertheless, as indicated in [7], many of the reviews indeed correspond to actual purchases. Also, as will be discussed later, the estimated product lifetime is used for dividing reviewers into different groups. Hence, it is reasonable to estimate a product's lifetime by its review time span. In our current work, we are only interested in products with complete lifetime, i.e., complete review time spans. 2.2 Early Reviewer Identification Given a complete product lifetime, we study how to divide the product lifetime into different stages so as to identify early reviewers. In the e-commerce website, the review posting process of users can be viewed as an adoption process of innovations. The process of adoption over time is typically illustrated as a classical normal distribution or "bell curve" and is divided into five stages [8]. Users are then categorized accordingly into five different groups, called innovators, early adopters, early majority, late majority and laggards. Following [8], [11], we apply the classic Rogers' bell curve theory to divide the product lifetime into five consecutive stages. In our datasets, the number of innovators is usually very small, and hence we combine innovators and early adopters as the early reviewers. In addition, we also combine early majority and late majority as majority, since it is usually difficult to reliably distinguish these two groups. Also, we transform the original intervals expressed in terms of the number of standard deviations from the mean into probabilities using simple cumulative distribution computation. The probability ranges for early, majority and laggards are [0, 0.16), [0.16, 0.84) and [0.84, 1] respectively.

3 System Study

Previous studies have highly emphasized the phenomenon that individuals are strongly influenced by the decisions of others, which can be explained by herd behavior. The influence of early reviews on subsequent purchase can be understood as a special case of herding effect. Early reviews contain important product evaluations from previous adopters, which are valuable reference resources for subsequent purchase decisions. As shown in, when consumers use the product evaluations of others to estimate product quality on the Internet, herd behavior occurs in the online shopping process. Different from existing studies on herd behavior, we focus on quantitatively analyzing the overall characteristics of early reviewers using large-scale real-world datasets. In addition, we formalize the early reviewer prediction task as a competition problem and propose a novel embedding based ranking approach to this task. To our knowledge, the task of early reviewer prediction itself has received very little attention in the literature. Our contributions are summarized as follows. We present a first study to characterize early reviewers on an e-commerce website using two real-world large datasets. We quantitatively analyze the characteristics of early reviewers and their impact on product popularity. Our empirical analysis provides support to a series of theoretical conclusions from the sociology and economics. We view review posting process as a multiplayer competition game and develop an embedding-based ranking model for the prediction of early reviewers. Our model can deal with the cold-start problem by incorporating side information of products. Extensive experiments on two real-world large datasets, i.e., Amazon and Yelp have demonstrated the effectiveness of our approach for the prediction of early reviewers.To predict early reviewers, we propose a novel approach by viewing review posting process as a multiplayer competition game. Only the most competitive users can become the early reviewer's w.r.t. to a product. The competition process can be further decomposed into multiple pairwise comparisons between two players. In a two-player competition, the winner will beat the loser with an earlier timestamp. Inspired by the recent progress in distributed representation learning, we propose to use a margin-based embedding model by first mapping both users and products into the same embedding space, and then determining the order of a pair of users given a product based on their respective distance to the product representation.

4 Implementation

There are three modules can be divided here for this project they are listed as below

- Upload products
- Product Review Based Order
- Rating and Reviews
- Data Analysis

From the above three modules, project is implemented. Bag of discriminative words are achieved

Upload Products

Uploading the products is done by admin. Authorized person is uploading the new arrivals to system that are listed to users. Product can be uploaded with its attributes such as brand, color, and all other details of warranty. The uploaded products are able to block or unblock by users.

Product Review Based Order

The suggestion to user's view of products is listed based on the review by user and rating to particular item. Naïve bayes algorithm is used in this project to develop the whether the sentiment of given review is positive or negative. Based on the output of algorithm suggestion to users is given. The algorithm is applied and lists the products in user side based on the positive and negative.

Ratings And Reviews

Ratings and reviews are main concept of the project in order to find effective product marketing. The main aim of the project is to get the user reviews based on how they purchased or whether they purchased or not. The major find out of the project is when they give the ratings and how effective it is. And this will helpful for the users who are willing to buy the same kind of product.

Data Analysis

The main part of the project is to analysis the ratings and reviews that are given by the user. The products can be analysis based on the numbers which are given by user. The user data analysis of the data can be done by charts format. The graphs may vary like pie chart, bar chart or some other charts.

5 Conclusion

In this paper, we have studied the novel task of early reviewer characterization and prediction on two real-world online review datasets. Our empirical analysis strengthens a series of theoretical conclusions from sociology and economics. We found that (1) an early reviewer tends to assign a higher average rating score; and (2) an early reviewer tends to post more helpful reviews. Our experiments also indicate that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity at a later stage. We have adopted a competition-based viewpoint to model the review posting process, and developed a margin based embedding ranking model (MERM) for predicting early reviewers in a cold-start setting.

6 References

[1] J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews," in WWW, 2016, pp. 625–635.

[2] N. V. Nielsen, “E-commerce: Evolution or revolution in the fastmoving consumer goods world,” nngroup. com, 2014.

[3] W. D. J. Salganik M J, Dodds P S, “Experimental study of inequality and unpredictability in an artificial cultural market,” in ASONAM, 2016, pp. 529–532.

[4] R. Peres, E. Muller, and V. Mahajan, “Innovation diffusion and new product growth models: A critical review and research directions,” International Journal of Research in Marketing, vol. 27, no. 2, pp. 91 – 106, 2010.

[5] L. A. Fourt and J. W. Woodlock, “Early prediction of market success for new grocery products.” Journal of Marketing, vol. 25, no. 2, pp. 31 – 38, 1960.

[6] B. W. O, “Reference group influence on product and brand purchase decisions,” Journal of Consumer Research, vol. 9, pp. 183–194, 1982.

[7] J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, “Imagebased recommendations on styles and substitutes,” in SIGIR, 2015, pp. 43–52.

[8] E. M.Rogers, Diffusion of Innovations. New York: The Rise of HighTechnology Culture, 1983.

[9] K. Sarkar and H. Sundaram, “How do we find early adopters who will guide a resource constrained network towards a desired distribution of behaviors?” in CoRR, 2013, p. 1303.

[10] D. Imamori and K. Tajima, “Predicting popularity of twitter accounts through the discovery of link-propagating early adopters,” in CoRR, 2015, p. 1512.

[11] X. Rong and Q. Mei, “Diffusion of innovations revisited: from social network to innovation network,” in CIKM, 2013, pp. 499– 508.

[12] I. Mele, F. Bonchi, and A. Gionis, “The early-adopter graph and its application to web-page recommendation,” in CIKM, 2012, pp. 1682–1686.

[13] Y.-F. Chen, “Herd behavior in purchasing books online,” Computers in Human Behavior, vol. 24(5), pp. 1977–1992, 2008.

[14] Banerjee, “A simple model of herd behaviour,” Quarterly Journal of Economics, vol. 107, pp. 797–817, 1992.

[15] A. S. E, “Studies of independence and conformity: I. a minority of one against a unanimous majority,” Psychological monographs: General and applied, vol. 70(9), p. 1, 1956.

[16] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in ICLR, 2013.

[17] A. Bordes, N. Usunier, A. Garc´ıa-Duran, J. Weston, and ´ O. Yakhnenko, “Translating embeddings for modeling multirelational data,” in NIPS, 2013, pp. 2787–2795.

[18] A. S. E, “Studies of independence and conformity: I. a minority of one against a unanimous majority,” Psychological monographs: General and applied, vol. 70(9), p. 1, 1956.

[19] M. L. S. D. X. W. L. S. Mingliang Chen, Qingguo Ma, “The neural and psychological basis of herding in purchasing books online: an event-related potential study,” Cyberpsychology, Behavior, and Social Networking, vol. 13(3), pp. 321–328, 2010.

Authors Profile

Sk. Heena, M.Tech., working as an Asst.Professor in the Department of Computer Science & Engineering in QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh, India.

Y.Beulah, pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

M.Syama Sindhuja pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

K.Maheswari,pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

V.Meghana,pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.