

CHARACTERIZING AND PREDICTING EARLY REVIEWERS FOR EFFECTIVE PRODUCT MARKETING ON ECOMMERCE WEBSITES

SAATHWIK CHANDAN NUNE
IIM KOZHIKODE

ABSTRACT: Online reviews have become an important source of instruction for users before manufacture an informed procure decision. Early reviews of a product tend to have a high effect on the ensuing 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 ecommerce platforms, i.e., Amazon and Yelp. In specific, we divide product lifetime into three uninterrupted phase, namely early, majority and straggler. A user who has posted a review in the early stage is contemplating as an untimely observer. 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 observer tends to assign a higher average rating score; and (2) an early observer 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 present a novel margin-based embedding model for early reviewer divination. Extensive experiments on two different ecommerce datasets have shown that our proposed approach outperforms a number of aggressive baselines.

Index-Terms:Early reviewer, Early review, Embedding model.

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 importance to monitor and attract at the early stimulation phase of a company. The pivotal role of early reviews has attracted extensive attention from marketing professional to convince consumer purchase neutral. For example, Amazon, one of the largest e-commerce company in the world, has advocated the Early Reviewer Program, 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. Based on the above conversation, we can see that early reviewers are especially important for product marketing. Thus, in this paper, we take the originality to study the department characteristics of early reviewers through their posted reviews on illustrative e-commerce platforms, e.g., Amazon and Yelp. We aim to conduct effective analysis and make accurate prognostication on early reviewers. This problem is strongly related to the adoption of innovations. In a generalized view, review posting process can be considered as an adoption of innovations, which is a theory that seeks to explain how, why, and at what rate new ideas and technology spread. The analysis and detection of early adopters in the diffusion of innovations have attracted much attention from the research community. Three fundamental elements of a diffusion process have been studied: attributes of an innovation, communication channels, and social network structures. However, most of these studies are theoretical examination at the macro level and there is a lack of quantitative explorations. With the rapid growth of online social platforms and the availability of a high volume of social networking data, studies of the diffusion of innovations have been widely conducted on social networks. However, in many application domains, social networking links or communication channel are unobserved. Hence, existing methods relying on social network structures or communication channels are not suitable in our current problem of predicting early reviewers from online reviews. To model the department of early reviewers, we develop a upstanding way to indicate the assumption process in two real-world large review datasets, i.e., Amazon and Yelp. More specially, given a product, the reviewers are sorted according to their timestamps for publishing their reviews. Following, we divide the product lifetime into three consecutive stages, namely early, majority and laggards. A user who has posted a review in the early phase is considered as an early commentator. In our work here, we mainly focus on two tasks, the first task is to analyze the overall characteristics of early reviewers compared with the majority and laggard reviewers. We characterize their rating behaviors and the helpfulness scores received from others and the correlation of their reviews with product popularity. The second task is to learn a forecast model which forecast early reviewers given a product. To analyze the characteristics of early reviewers, we take two important metrics associated with their reviews, i.e., their review ratings and helpfulness scores assigned by others. We have found that (1) an early reviewer tends to assign a higher average rating score to products; and (2) an early reviewer tends to post more helpful reviews. Our above findings can find relevance in the classic principles of personality variables theory from social science, which mainly studies how innovation is spread over time among the participants:

- (1) earlier adopters have a more favorable attitude toward changes than later adopters; and
- (2) earlier adopters have a higher degree of opinion leadership than later adopters.

We can relate our findings with the personality variables theory as follows: higher average rating scores can be considered as the favorable attitude towards the products, and higher helpfulness votes of early reviews given by others can be viewed as a proxy estimate of the perspective leadership. Our analysis also indicates that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity. We further explain this

finding with the herd behavior widely studied in economics and sociology. Herd behavior refers to the fact that individuals are strongly influenced by the decisions of others

LITERATURE REVIEW

- Ting Bai, Jian-Yun Nie[1] provided a 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. In viewing review posting procedure as a multiplayer competition game, we propose a novel margin based embedding model for early reviewer forecast. Experimenting on two different e-commerce datasets have shown that our proposed system outperforms a number of competitive baselines
- Julian McAuley, Alex Yang[2] Provided a Online audits are regularly our first port of call while considering items and buys on the web. While assessing a potential buy, we may have a particular inquiry as a main priority. To answer such inquiries we should either swim through colossal volumes of buyer audits planning to discover one that is pertinent, or generally suggest our conversation starter straightforwardly to the network by means of a Q/A framework. In this paper we would like to meld these two ideal models: given a huge volume of beforehand addressed questions about items, we trust to consequently realize whether an audit of an item issignificant to a given question. We define this as a machine learning issue utilizing a blend of-specialists compose system—here each audit is a 'specialist' that gets the opportunity to vote on the reaction to a specific question; all the while we take in an importance capacity with the end goal that 'applicable' audits are those that vote accurately. At test time this scholarly importance work enables us to surface audits that are important to new questions on request.
- Matthew J. Salganik, Peter Sheridan Dodds, Duncan J. Watts [3] provided Collaborative filtering has proven to be valuable for recommending items in many different domains. Here, we explore the use of collaborative filtering to recommend research papers, using the citation web between papers to create the ratings matrix. We tested the ability of collaborative filtering to recommend citations that would be suitable for additional references to target a research paper. We analyzed six methods for selecting citations, evaluating this through offline demonstration against a database of over 186,000 research papers hold in Research Index. We also performed an online demonstrate with over 120 users to measure user opinion of the effectiveness of the algorithms and of the utility of such recommendations for common research tasks. We came across large differences in the accuracy of the algorithms in the offline experiment, especially when balanced for coverage. In the online experiment, users felt they received quality recommendations, and were enthusiastic about the idea of receiving recommendations in this domain.
- Julian McAuley, Christopher Targett, Qinfeng ('Javen') Shi, Anton van den Hengel[4] intrigued here in revealing connections between the appearances of sets of items, and especially in displaying the human idea of which objects supplement each other and which may be viewed as satisfactory options. We accordingly try to

demonstrate what is an on a very basic level human idea of the visual connection between a couple of articles, as opposed to just displaying the visual similitude between them. There has been some enthusiasm generally in displaying the visual style of spots, and objects. We, interestingly, are not looking to show the individual appearances of objects, yet rather how the presence of one question may impact the attractive visual characteristics of another.

EXISTING SYSTEM:

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.

PROPOSED SYSTEM:

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.

A. Frequency based Itemset Mining Regular itemset mining is a conventional and significant problem in data mining. An itemset is repeated if its support is not less than a brink stated by users. Conventional regular itemset mining approaches have chiefly regarded as the crisis of mining static operation databases. In the operation data set regular itemsets are

the itemsets that happen often. To recognize all the regular itemsets in a operation dataset is the objective of Frequent Itemset Mining. Within the finding of relationship rules it created as a phase, but has been simplified autonomous of these to several other samples. It is confronting to enlarge scalable methods for mining regular itemsets in a huge operation database as there are frequently a great number of diverse single items in a distinctive transaction database, and their groupings may form a very vast number of itemsets.

B. Utility based Itemset Mining By seeing the circumstance of usage as précised by the user a high utility itemset is the one with utility value larger than the minimum brink utility. A wide topic that wraps all features of economic utility in data mining is known to be utility-based data mining. It includes the work in cost-sensitive education and dynamic learning as well as work on the recognition of uncommon events of high effectiveness value by itself. By maintaining this in mind, we at this point offer a set of algorithms for mining all sorts of utility and frequency based itemsets from a trade business deal database which would considerably aid in inventory control and sales promotion. Consideration of a utility based mining approach was motivated by researchers due to the limitations of frequent or rare itemset mining, which permits a user to suitably communicate his or her views regarding the usefulness of itemsets as utility values and then find itemsets with high utility values higher than a threshold. Identifying the lively customers of each such type of itemset mined and rank them based on their total business value can be done by these set of algorithms. This would be enormously supportive in developing Customer Relationship Management (CRM) processes like campaign management and customer segmentation. In all types of utility factors like profit,significance, subjective interestingness, aesthetic value etc the utility based data mining is a newly absorbed research area. This can add economic and business utility to existing data mining processes and techniques. A research area inside utility based data mining identified as high utility itemset mining is intended to discover itemsets that introduce high utility.

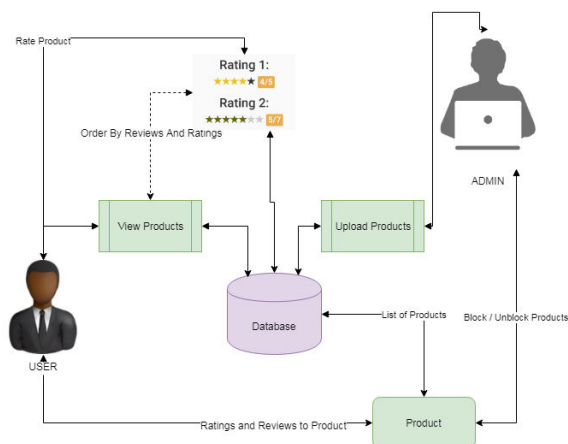
C. Correlation feature selection Feature selection is a preprocessing step to machine learning which is constructive in diminish dimensionality, detach immaterial data, increasing learning perfection, and improving result comprehensibility

1) Steps of feature selection A feature of a subset is good if it is highly correlated with the class but not much correlated with other features of the class. Steps: a. Subset generation: We have used four classifiers to rank all the characteristics of the data set. Then we have used top 3, 4, and 5 characteristics for classification Subset evaluation: Each classifier is applied to generated subset. Stopping criterion: Testing process continues until 5 characteristics of the subset are selected Result validation: We have used 10-fold cross acceptance method for testing each classifier's accuracy.

D. Classification techniquesNBTreeNBTree is a simple hybrid algorithm with Decision Tree and Naïve-Bayes. In this algorithm the smple concept of recursive partitioning of the schemes remains the same but here the difference is that the leaf nodes are naïve Bayes categorizers and will not have nodes predicting a single class. Naïve Bayes The Naïve Bayes classifier technique is used when dimensionality of the inputs is high. This is a easy algorithm but gives good output than others. We are using this to find the dropout of students by calculating the probability of each input for a predictable state. It trains the weighted training

data and also helps prevent over fitting Instance-based-k-nearest neighbor. In this technique a new item is divided by comparing the memorized data items using a distance measure. For this we require storing of a dataset. Matching of items is happened by putting them close to original item. Nearest neighbors can be happened by using cross-validation either automatically or manually.

ARCHITECTURE:



MODULES:

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

MODULE DESCRIPTION:

1. 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.

2. 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.

3. 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.

4. 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.

CONCLUSION We have studied the novel task of early reviewer characterization and prediction on two real-world online review datasets. Our actual analysis strengthens a series of theoretical conclusions from sociology and economics. We found that an early reviewer tends to assign a higher average rating score; and 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.

REFERENCES

- [1]. J. McAuley and A. Yang, "Addressing complex and subjective product-related queries with customer reviews," in WWW, 2016, 625–635.
- [2]. N. V. Nielsen, "E-commerce: Evolution or revolution in the fast- moving consumer goods world," nngroup. com, 2014.
- [3]. W. D. J. Salganik M J, Dodds P S, "Experimental study of in- equality and unpredictability in an artificial cultural market," in ASONAM, 2016, 529– 532.
- [4]. R. Peres, E. Muller, and V. Mahajan, "Innovation diffusion and new product growth models: A critical review and research direction," International Journal of Research in Marketing, 27(2), 2010, 91 – 106.
- [5]. L. A. Fourt and J. W. Woodlock, "Early prediction of market success for new grocery products." Journal of Marketing, 25(2), 1960, 31 – 38.
- [6]. B. W. O, "Reference group influence on product and brand purchase decisions," Journal of Consumer Research, 9, 1982, 183–194.
- [7]. J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image- based recommendations on styles and substitutes," in SIGIR, 2015,. 43–52.
- [8]. E. M.Rogers, Diffusion of Innovations. New York: The Rise of High- Technology 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, 1303.
- [10]. D. Imamori and K. Tajima, "Predicting popularity of twitter ac- counts through the discovery of link-propagating early adopters," in CoRR, 2015, 1512.
- [11]. X. Rong and Q. Mei, "Diffusion of innovations revisited: from social network to innovation network," in CIKM, 2013, 499– 508.
- [12]. I Mele, F. Bonchi, and A. Gionis, "The earlyadopter graph and its application to web- page recommendation," in CIKM, 2012, pp. 1682–1686.
- [13]. Y.-F. Chen, "Herd behavior in purchasing books online," Computer in Human Behavior, 24(5), 2008, 1977–1992.

- [14]. Banerjee, “A simple model of herd behaviour,” *Quarterly Journal of Economics*, 107, 1992, 797–817.
- [15]. A. S. E, “Studies of independence and conformity: I. a minority of one against a unanimous majority,” *Psychological monographs: General and applied*, 70(9), 1956,1.
- [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’ia-Dura’n, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi- relational data,” in *NIPS*, 2013,2787–2795.
- [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), 1956, 1.
- [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), 2010, 321–328.
- [20]. V. G. D. W. Shih-Lun Tseng, Shuya Lu, “The effect of herding behavior on online review voting participation,” in *AMCIS*, 2017.