IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 11, November. 2019, Series -I, PP 05-09

A Survey on Early Reviewers prediction techniques on E-Commerce Websites

Arifa.P[1], Ms. D. Priyadarshini[2]
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[1]
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu India[2]
Corresponding Author: Arifa.P
Received 19 October 2019; Accepted 5 November 2019

Abstract: E-commerce websites become very popular and tremendous growth in product marketing. Customer satisfaction and effective fulfillment of customer needs are more important in e-commerce applications. To achieves effective results and feedback in e-commerce applications, the need for effectively gathering user reviews and feedback. However, the review collection and analysis became very challenging due to its size and unique nature early review prediction from the posted reviews in real-time large scale e-commerce websites are more challenging. So, the utilization of data mining will resolve these issues. There are numerous approaches and techniques proposed to handle these issues. This survey gives a description and study on those techniques with various factors. Especially, Top-K algorithms and Hidden Markov model techniques are more prominent techniques in this filed. Finally, the survey gives a conclusion about the merits and demerits of such techniques and navigates to the future scope.

Index Terms-E-Commerce, Review Prediction, Data mining, Early reviewer, Early review, Embedding model, Top-K, Hidden Markov Model.

I. INTRODUCTION

Web based business, in other words it will be called as E-Commerce or Electronic business. In Online business all the transactions are going to be done though internet only. Users can easily get the desired products. The services, payments and guidance for the usage of the product will completely do by using web based technologies. The online business is same as the ordinary business, But the only difference between is in Business all the transactions are done by web based technology only. In online we can get all the services like banking, movie tickets, hotel booking, air tickets, E-booking, trading, etc. In online we can get any type of product .There are several websites for online business example Amazon, Flipkart, Paytm, Snapdeal, etc. Each website is having a numerous type of product varieties .For example, Amazon website is one of the greatest website in E-commerce, firstly it started an online book store with a wide variety of books later it became a store for all the products. Today in India the average sale of products only from the amazon is about 18 million products. For such online marketing, huge numbers of reviews are given by the users for the products they purchased from the site. Based on those reviews the other users can able to know what is the good product. Such a user comments are having a high knowledge on the product. These reviews are very much important for both the consumer and the firms. As the consumer can able to know the quality of the product, whereas the firm can able to get the feedback of the product. So the firm can update the product according to the consumer's requirement and they can get improve in online marketing, development of product and in maintaining the relationship with the consumer.

The sound and speed of online reviews are noted by purely visiting e-commerce and customer rating sites, such as Yelp and Amazon. There is better variation across the possible industry portion that is distinct from others for reviews (such as hotels, restaurants, e-commerce, home services, etc.). In web based product reviews which are posted by customers who have previously purchased products help future customers make better purchase decisions. Reviews can be defined as peer-generated, open-ended comments about the product posted on company or third party websites. Since reviews are autonomously updated by customers themselves without corporate efforts, they are perceived as a sustainable form of word of mouth in e-business. Therefore, as the reviews gather, it becomes roughly impossible for customers to study all of them; In addition, poorly authored low-quality reviews can even cause inconvenience. Thus, it becomes essential for online business companies to recognize helpful reviews and selectively present them to their customers. Along with the large number of languages that reviews are written in accuracy is a problem with online reviews, since the very great

extent majority of reviews are without a label, which means it is not easily known whether the review is fake or not. As an extra factor, standard machine learning algorithms tend to separate into pieces as a result of a blow down and become not producing any significant when dealing with data of this size, which poses a problem when trying to make a formal application.

The majority of the people will read before purchasing of the product. Based on the review the users can purchase the products for online. This approach describes to find early reviewers such as who are all post reviews on initial stage of the product. 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. On behalf of this reason, early reviewers develop into the emphasis to monitor and attract at the early promotion period of an industry or a company. The essential position of early reviews has attracted broad attention from advertising practitioners to induce consumer purchase intentions. For example, Amazon, one of the largest e-commerce companies 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. A further associated program, Amazon invites the most believable reviewers on Amazon to send opinions about new and pre-release items to help their fellow customers make informed purchase decisions. Based on the previous deliberations, peoples can see that early reviewers are extremely significant for product marketing. Thus, in this approach, take the idea to read the performance characteristics of early reviewers through their posted reviews on representative e-commerce platforms, e.g., Amazon and Yelp. In this approach our goal is to achieve to conduct effective analysis and make exact prediction 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.

II. LITERATURE REVIEW

In this paper [1] the author describes online business is the easiest way of shopping. In web based production, users can purchase the products by reading the feedbacks or reviews of the previous users who are used the products earlier. Based on people's opinions the product can get grade. But the user has to view many number of reviews for a particular product in order to purchase the better product. It was the time taking process. In this approach we are invented to suggest a system that we can straightly collect the reviews of the products from online and by comparing those reviews we can receive the best items based on the positive opinions given by previous users of that product.

In this paper [2] the author describes In today's developed world, every minute, people around the globe express themselves via various platforms on the Web. And in every minute, a mass amount of unstructured data is generated. This data is in the form of text which is gathered from forums, social media websites, and reviews. Such data is termed as big data. User opinions are related to a wide range of topics like on particular products also. These reviews can be mined using various technologies and are of at most importance to make predictions since they directly convey the viewpoint of the masses. Online reviews also have become an important source of information for users before making an informed purchase decision. Early reviewer's ratings and their received helpfulness scores are likely to influence product popularity. The challenge is to gather all the reviews, also calculate and analyze the ratings, in order to find a refined product, that scores high rating.

In this paper [3] the author describes, the percentage of purchasing products by the user has been increased drastically through web. Customers even have the ability of sharing their thoughts about the exact product on web in the form of reviews, online blogs, online comments etc. Lot of users view and find review information given on web to take decisions for purchasing products. Some users may provide the reviews for often exaggerating its benefits the sale of the product or to decrease the sale. This may puzzle the consumers who rely on the reviews to purchase a product. So, there is a need to find the honest reviews and remove fake reviews that are added by malicious or fraud user. The proposed system comes up with the solution for this problem. Leading events has been used to find the time interval between the reviews. The presented system mines the live periods such as most important sessions to precisely set the hierarchical fraud. These important sessions can be useful for finding the local anomaly instead of international anomaly of product reviews. After this to recognize the rating, reviews and hierarchy of the product we observe three facts, they are online rating based facts, online review based facts and hierarchy facts. In addition, we propose an optimization-based aggregation schemes to combine all the facts for fraud detection. The predictions of this optimization are done on synthetic dataset that are gathered. The classified and described product review sequence helps web users to recognize review contents easily in a short time.

In this paper [4] the author describes online customer reviews are a sustainable form of word of mouth (WOM) which play an increasingly important role in e-commerce. Through, low quality reviews can often cause problems to review viewers. The reason of this approach is to automatically predict the usefulness of reviews. This approach analyzes the individuality embedded in product reviews across five various product types and explores their effects on review usefulness. In addition, four data mining techniques were examined to describe the one that best predicts review usefulness for each product type using five real-life review datasets gathered from Amazon.com. The conclusion shows that reviews for various product types have various psychological and linguistic personalities and the factors affecting the review usefulness of them are also different. Our solutions also indicate that the support vector regression techniques to predicts review usefulness most perfectly among the four schemes for all five datasets. This approach contributes to improving efficient utilization of online reviews.

In this paper [5] the author describes Product reviews are posted online by the hundreds and thousands for popular products. Managing such a high volume of frequently generated online content is a challenging task for purchaser, sellers and researchers. The reason of this approach is to rank the vast number of reviews using their find useful scores. The useful score is predicted using attributes extracted from review contents, product description, and consumer query-respond data of a product using the random-forest classifier and gradient boosting regressed. The approach classifies reviews into low or high quality with the random-forest classifier. The usefulness scores of the high-quality reviews are merely predicted using the slope boosting regressed. The usefulness scores of the small-quality reviews are not measured because they are no way going to be in the top k reviews. They are immediately added at the end of the review list to the review-listing website.

In this paper [6] the author describes as the popularity of free-form user generated reviews in e-commerce and review websites continues to increase, there is a growing need for automatic techniques that separate through the vast number of reviews and recognize quality content. Online review usefulness modeling and recognition is a task which studies the factors that describe review uses and attempts to perfectly predict it. This survey paper gives an overview of the most related work on product review uses prediction and accepting in the past decade, discusses gathered insights, and gives guidelines for prospect research.

In this paper [7] the author describes online review is an important form of electronic word of mouth (eWOM) that helps customers make purchasing decisions. In a set of reviews, the review with the most helpfulness votes is seen as most helpful. While researchers have demonstrated how review and reviewer characteristics influence helpfulness votes, a largely uninvestigated issue is how herding behaviors can influence customers' voting participation and direction. Drawing on herd behavior literature, we propose that review voters will discount their own information when faced with clear and strong signals from previous voters. Thus, they will herd previous voters' voting direction. On the other hand, review voters will value their own judgments when faced with weak signals from previous voters. Herding can influence both a voter's perception of a review's helpfulness and his/her vote. This research extends review helpfulness literature that herd behaviors could moderate customers' perception of review helpfulness and voting direction.

In this paper [8] the author describes online reviews provided by consumers are a valuable asset for e-Commerce platforms, influencing potential consumers in making purchasing decisions. Therefore, these reviews are of unreliable quality, with the useful ones covered deep within a heap of non-informative reviews. In this approach, we attempt to routinely recognize review quality in terms of its uses to the end customers. In difference to existing works in this domain exploiting a variety of syntactic and community-level attributes, we look deep into the semantics of reviews as to what makes them helpful, giving interpretable explanation for the same. We recognize a set of reliability and semantic factors, all from the text, ratings, and timestamps of user-created reviews, making our system common across all communities and domains. We determine review features in terms of a variety of latent factors like the expertise of its author, his decision about the fine-grained facets of the underlying product, and his writing style. These are direct into a Hidden Markov Model – Latent Dirichlet Allocation based model to combine infer: (i) online reviewer expertise, (ii) online item facets, and (iii) online review helpfulness. Large-scale tasks on five real-world datasets from Amazon display major development over state-of-the-art baselines in predicting and ranking useful reviews.

In this paper [9] the author describes Online reviews are often our first port of call when considering products and purchases online. In this approach we expect to fuse these two paradigms: given a large volume of existing solution queries about products, we hope to robotically learn whether a review of a product is applicable to a given query. We prepare this as a machine learning problem using a combination-of-experts-type framework—here each review is an 'expert' that gets to vote on the answer to a exact query; At the same time we learn a significance function such that 'relevant' reviews are those that vote perfectly. At checking time this studied relevance function allows us to outside reviews that are relevant to new queries on-required. We measured our technique, Moqa, on a novel corpus of 1.4 million questions (and answers) and 13 million

reviews. We display more quantity that it is efficient at finding both binary and open-ended questions, and qualitatively that it surfaces reviews that human calculates consider to be relevant.

In this paper [10] the author describes, this study uses eye-tracking method to investigate consumers' online review search behavior by suggesting that it needs to consider the type of product reviewed. A review-product congruity scheme was testified through a self-report survey and eye-tracking tasks. The proposition states that consumers of search products expect to seek attribute based reviews, while consumers shopping for experience products tend to seek experience based reviews. The conclusion of our initial empirical task support our hypotheses by displaying consumers' more energetic and positive responses to features based reviews when shopping for search products and to experience based reviews when purchasing experience products. The other task was conducted with eye tracking technique to gain additional insights. We found that customers of search products are attracted and occupied more deeply by attribute based reviews. In addition, when they look familiar items, the difference of their fixations on experience based online reviews and quality based reviews is important, and thus the proposition is partially supported. This approach explained our present consideration of customers online review search performance by subsuming product type, which is essential and helpful, and gives references on the categorization and presentation of reviews to help customers' product finding and decision making. Furthermore, evaluation of traditional empirical scheme and eye-tracking method can help deepen our supportive of complex customer online shopping behavior.

In this paper [11] the author describes, 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.

In this paper [12] the author describes, provided approach for concept Due to the dynamicity, new well known records consistently show up and vanish in miniaturized scale blogging administrations. Early identification of new records that will wind up mainstream in future is an essential issue that has a few applications, for example, slants location, viral showcasing, and client suggestion. Estimation of prominence of a record is additionally valuable for approximating the nature of data it posts. Estimation of the nature of data is vital in numerous applications, yet it is for the most part hard to gauge it without human mediation. Comparative thought has additionally been effectively connected to small scale web journals with connecting capacities. These certainties demonstrated that there is high relationship between the notoriety and the nature of data. In this manner, the estimation of forthcoming notoriety of new records, which have not yet settled the prevalence they merit, is additionally helpful for estimation of the quality.

In this paper [13] the author describes, online business is the easiest way of shopping. In ecommerce business, customers can buy the items by reading the feedbacks or reviews of the other customers who are used the items earlier. Based on customer's opinions the product can get grade. But the user has to view a lot of reviews for a single product in order to get the best product. It was the lot of time taking procedure. In this approach we are believed to propose a method that we can straightly collect the reviews of the products from online and by comparing those reviews we can get the best product based on the positive opinions given by previous users of that product.

III. CONCLUSION

From this paper, Its found that the early detection of E-commerce website reviewers is mandatory for effective product marketing. Thus, in this paper, a detailed study of the early reviewer's prediction is performed. This paper summarizes the challenges, issues, and techniques proposed by various authors in the literature. The behavior analyses of early reviewers through their past activites such as posted reviews on representative e-commerce platforms are creating a serious data mining problem. So finding appropriate data mining techniques to handle such problems is more necessary. Before inventing a new technique, there is a need to study the existing techniques and approaches on the same issue. This survey gives a summary of the problem, which is strongly related to the adoption of innovations. The opinions on the product and its services in the E-commerce websites should be handled properly. From this survey, a need for innovative technology with all metrics should be developed. This survey concludes that the algorithms should adopt for all real-time E-commerce websites.

REFERENCES

[1]. Sayli A. Hon, Pritee V. Shirsath, Sheetal K. Pendhari, "Opinion Mining and Predicting Reviews for E-commerce ", International Journal of Scientific Research and Engineering Development-– Volume 2 Issue 3, May –June 2019.

[2]. D. Vetriselvi, D. Monisha, M. Monisha varshini, "Predicting Review Ratings for Product Marketing", International Research Journal of Engineering and Technology, Volume: 06 Issue: 03 Mar 2019.
[3]. F. Femila, S. Janakipriya, R. B. Nivetha Sruthi, S. Rohini, "Predicting Early Reviews for Effective Product Marketing on E-Commerce Websites", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Issue 2, Volume 5, 2019.
[4]. Yoon-Joo Park, "Predicting the Helpfulness of Online Customer Reviews across Different Product Types", Seoul National University of Science and Technology, 2018.
[5]. SunilSaumyaa, Jyoti Prakash Singh, "Ranking online consumer reviews", Electronic Commerce Research and Applications, 2018.
[6]. Gerardo Ocampo Diaz and Vincent Ng, "Modeling and Prediction of Online Product Review Helpfulness: A Survey", Association for Computational Linguistics, 2018.
[7]. V. G. D. W. Shih-Lun Tseng, Shuya Lu, "The effect of herding behavior on online review voting participation," in AMCIS, 2017.
[8]. Subhabrata Mukherjee, Kashyap Popat, Gerhard Weikum, "Exploring Latent Semantic Factors to Find Useful Product Reviews, Society for Industrial and Applied Mathmatics, 2017.
[9]. Julian McAuley, Alex Yang, "Addressing Complex and Subjective Product-Related Queries with Customer Reviews", International World Wide Web Conference Committee, 2016.
[10]. Jing Luan, Zhong Yao, FuTao Zhao, Hao Liu, "Search product and experience Product online reviews: An eye tracking study on consumers' review search behavior", Computers in Human Behavior, 2016.
[11]. J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image based recommendations on styles and substitutes," in SIGIR, pp. 43–52,2015.
[12]. D. Imamori and K. Tajima, "Predicting popularity of twitter accounts through the discovery of link-propagating early adopters," in CoRR, 2015.
[13]. Pusarla.Divya Bharathi , A.Gauthami Latha , "Predicting User Opinion in E-commerce Website Using Aggregate Ranking Algorithm", International Journal of Innovative Research in Science, Engineering and Technology, 2015.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Arifa.P." A Survey on Early Reviewers prediction techniques on E-Commerce Websites." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 11, 2019, pp. 05-09.