IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Predicting Early Reviews for Product Marketing On E-Commerce Site

[1] Cincy Raju; [2] Deepa Rajan S; [3] Dr. Radhakrishnan B

[1] CSE Department BMCE,
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

[2] CSE Department, BMCE,
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

[3] CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

Abstract- Online review has become an important source of information for user before purchasing the product. Customer reviews are a form of customer feedback on electronic commerce and online shopping sites. The product lifetime is divided into three consecutive stages namely, early, majority and laggards. Product reviews contain useful opinion, comments and feedback towards their product. In the proposed work early reviews of product are using a machine learning algorithm followed by frequency based ranking and correlation based ranking. In Frequency based ranking, that extracts the positive review or positive feedback about the particular product and to give a frequency score. Based on frequency score and using correlation based ranking, to find the product sale. Finally, Aggregate positive review response and negative review response from the above ranking method to get a feedback. Feedback means, whether the product is trendy or non-trendy.

Keywords- *Machine learning, Early reviewer, Early reviews, Multilayer perceptron*

1. Introduction

The arrivals of e-commerce platform for online shopping allows the buyer to publish and share their opinion after the purchase of a product and sharing their experience by posting product reviews. Recently most of the buyers refer to online product reviews. The purchase of the product mostly depends on the review and tends to have high impact on sales. The posted reviews plays more important role to the buyer, reading these reviews give a detailed description about product. The person who posts a brief review in the early stage is known as early reviewer. The success and failure of a new product or service can be determined by the posted reviews and their opinions. To identify the early reviewers and their feedback can help companies to adjust their marketing strategies and improving product design, which will gradually lead to the success of their new product .Buyer decide which product to buy based on the positive posted review and provide advantage to the buyer in taking decisions on purchasing the product. Mining of the customer reviews will involve automate extraction of reviews and ratings, cleaning the data, quantitatively analyzing the ratings, qualitatively analyzing the reviews through opinion mining or sentiment analysis and arriving at a score for a specific product that will help customer differentiate several products based on customer reviews.

The customer reviews are extracted from Amazon dataset and the features star ratings, text reviews, helpfulness score of a review, product feature spoken about in the review and the date of the review are considered. We used Natural Language Processing techniques to determine the polarity of the reviews and also arrived at a score for a specific product by including:

- Star Rating
- Number of Positive Reviews
- Number of Negative Reviews
- Helpfulness score of reviews
- Age of Review

Used these scores to compare two or more products and recommend the best product to the customer.

As the number of products being sold online increases, it is becoming increasingly difficult for customers to make purchasing decisions based on only pictures and short product descriptions. On the other hand, customer reviews, particularly the text describing the features, comparisons and experiences of using a particular product provide a rich

source of information to compare products and make purchasing decisions.

Recently online shopping marketing becoming more popular due to its rapid sales and the numbers of customer reviews that make a product grow rapidly.

2. Related Work

Online customer reviews are helping consumers to decide which products to buy and also the companies to understand the buying behavior of consumers and to recommend compare of products sold online. Natural language processer is used to automatically read reviews and Naive Bayes is used for classification, ie: to determine the polarity of reviews. Then the reviews of the product are extracted from the polarity of those features. So the customer can rate the better of two products based on various criteria including the star ratings, date of review, the helpfulness score of the review and the polarity of reviews.[1]

The consumers naturally gravitate to reading reviews in order to decide whether to buy a product. However, the high volume of reviews that are typically published for a single product makes it harder for individuals to locate the best reviews and understand the true underlying quality of a product based on the reviews. There are two ranking mechanisms for ranking product reviews: a consumer-oriented ranking mechanism ranks the reviews according to their expected helpfulness, and a manufacturer Oriented ranking mechanism ranks the reviews according to their expected product on sales. The subjectivity analysis can give useful clues about the helpfulness of a review and about its impact on sales.[2]

A customer reviews are not only helpful for potential customers, but also are helpful for the manufacturers of the products to raise the level of their products and services. Opinion Mining is playing a major role to summarize customer reviews and make it easy for online customers to determine whether to purchase the products or not. *K-NN* is used for the review sentences' classification into two classes (subjective, objective).[3]

The opinion mining aims at summarizing the content of reviews for a specific brand, Product, or manufacturer. The task to rank products are based on sentiment information. This includes

1) The identification of gold rankings which acts as a fundamental for an objective function and evaluation.
2) Methods to rank products based on review information.

As baseline methods, we use the average star ratings and review frequencies. In addition, aspect-specific rankings can be used to measure the impact of specific aspects on its ranking.[4]

The e-commerce is increasingly becoming popular; the number of customer reviews for a product receives grows rapidly. However, for popular products, many online product reviews exist but for other products, reviews are very few. These online discussions about particular products may help other online users to make a decision in buying/ not buying those products, like in amazon.com1 and ebay.com2. Since an enormous number of unstructured and ungrammatical reviews on a product exist, opinion mining is getting a crucial research area for better decision making of buying products.

For applying opinion mining approach to summarize the unstructured and ungrammatical users' reviews, based on Support vector machine (SVM), two levels of classification are applied:
1) Features classification
2) Polarity classification for every feature class.[5]

3. Proposed Approach

3.1 Data Preparation

The details are taken from Amazon site datasets. Datasets are in the form of comma separated values. Each product review is a textual comment posted by a buyer of a product and is accompanied by its publish time stamp. A review is associated with a rating score in a five-star scale. By giving review for the product users can also vote the helpfulness score using a binary choice of Yes or No button.

3.2 Data Preprocessing

With the help of NLP we first removed reviews from anonymous users, to associate each review with a unique product buyer and then remove duplicate reviews by multiple versions of the same product. Also remove inactive users and unpopular products. With Tokenization the unwanted words or the word which cannot be evaluated like e.g. is, am, are etc. are being removed from the product review.

The Product's review is being collected from online shopping site's. Reviews are divided into three times based stages are as Early, Majority and Laggards. The early reviewer of a product refers to who has posted the review in the early stage. There are two ranking algorithms such as frequency based ranking and correlation based ranking is

applied to the early reviews. Reviews are extracted from e-commerce sites and then tokenization is done. The words are classified into two classes such as positive word and negative word. The review feature is compared to corresponding classes by using 'NLP' process, NLP process compared and gives the details of feature belong which class. In frequency based ranking, review for a particular product to give a number of positive and negative feedback based on 'like and comments'. The count of likes and comments give the tend frequency score details. Correlation based ranking is an association between frequency and product trending. Based on frequency score and using correlation based ranking are to find the product sale. Finally pooling positive review feedback and negative feedback based on early reviews. Using multilayer perceptron we can predict whether the product is trendy or not.

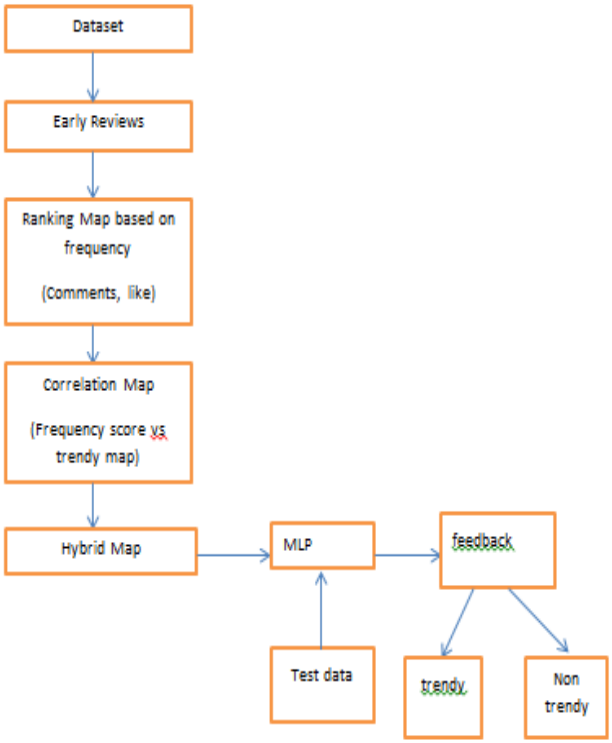

Fig.1 Proposed system

3.3 Multilayer Perceptron

In a multilayer perceptron, the neurons are arranged in an input layer, an output layer and one or more hidden layers. The MLP algorithm is a very good algorithm to use for the regression and mapping. It can be used to map an *N*-dimensional input signal to an *M*-dimensional output signal. This mapping can also be non-linear.

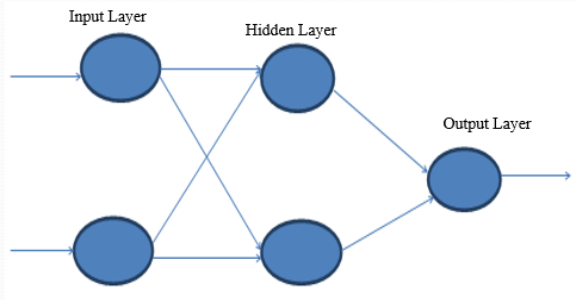

Fig 2: The Multilayer Perceptron Model

4. Result Analysis and Discussion

Recent results of Amazon dataset, are divided as test data at 20% and the remaining 80% as training data. The multilayer perceptron technique is applied. The approach has been tested with dataset of 535 sentences; the accuracy of the proposed system is 94%.

Table 1: Result Analysis

Algorithm	Accuracy
Multilayer perceptron	94%
Svm	86%
Naïve Bayes	84%

5. Conclusion

The online review's for a product has become an important source of information and get the details about it from the buyer before purchasing or ordering the product. The Product's review is being collected from online shopping site's and reviews contain useful opinion, comments and feedback towards their product. The feedback is predicted using a multilayer perceptron followed by frequency based ranking and correlation based ranking. Feedback means whether the product is trendy or non-trendy.

References

[1] Venkata Rajeev P and SmrithiRekhaV, Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features.
[2] Anindya Ghoseand Panagiotis G. Ipeirotis, Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews.

[3] Jawad Khan And Byeong Soo Jeong, Summarizing Customer Review Based On Product Feature and Opinion.
[4] Wiltrud Kessler, Roman Klinger, and Jonas Kuhn, Towards Opinion Mining from Reviews for the Prediction of Product Rankings.
[5] Taysir Hassan A. Soliman and Mostafa A. Elmasry, Utilizing Support Vector Machines in Mining Online Customer Reviews.

Author Profile

Cincy Raju is pursuing her Mtech degree on CSE at APJ Abdul Kalam Technological University. Her research interest are in image processing, data mining, machine learning.

Deepa Rajan S is working as Assistant Professor in Computer Science and Engineering Department. She has 11 years' experience in teaching. Her research interests focuses on Data Security, Image Processing.

Dr.Radhakrishnan B is working as the Head of CSE department. He has more than 14 years, experience in teaching and has published papers on data mining and image processing. His research interests include image processing, data mining, image mining.