

Product Recommendation System using Opinion Mining

Ajith Kumar V^{#1}, Arun B^{#2}, Balamurugan J^{#3}, Nancy Deborah R^{#4}

^{1,2,3}Student, ⁴Assistant Professor, Department of Information Technology, Velammal College of Engineering and Technology, Madurai, Tamilnadu, India.

Abstract

In this technological era, people are drowned into the sea of internet. After food, clothes and shelter, the internet has become a basic need of every citizen. They can get whatever they want from the internet, like if they want to know something, Wikipedia educates them; if they want to see some videos, YouTube shows them; if they want to eat something, Zomato or Swiggy gets them. This has made people so lazy and they want to obtain everything from home itself. Now-a-days, people feel reluctant to go shopping and hence have started to order them online. The internet has become the biggest shopping stop. Before buying things, people go through the reviews and try to buy the best product. It is quite usual now, but, it is not guaranteed that they are buying the genuine product. Today, the product manufacturers are trying to fool people by submitting fake reviews and make the people fall for their products. These fake reviews can be monitored and the genuine reviews can be filtered out with the help of Machine Learning. The reviews can be analyzed by Opinion mining. When a person tries to submit two or more reviews for a single product via same IP address, it can be identified and can be removed. This system eliminates the fake reviews and recommends the best products for the people to buy.

Keywords - Opinion Mining, Machine Learning, IP address, fake review

I. INTRODUCTION

Today, the world has been taken over by the internet. In this fast-moving technological world, the internet is accessed by almost everyone who owns a smartphone or a desktop computer or a laptop. For example, students carry out their projects by referring to online resources and study through online courses, and one of the most common things happening on the internet is the online shopping. People prefer to buy products online rather than going out and buying at the actual shops. Before buying any product, people look at the ratings and reviews of that product to know whether the product is good or not. Seller selling products on the web often ask or take reviews from customers about the products that they have purchased. As e-commerce is growing day-by-day, the number of reviews received from customers about the product increases. So, people come across various

reviews in the website, but the people could not identify if the product is genuine. For a popular product, the reviews can go up to thousands. This creates difficulty for the potential customer to read them and to decide whether to buy or not buy the product. In some review websites, some good reviews are added by the product company people itself to make product famous. So, customers buying products online look at those reviews, think that the product is good and buy them. To find out and remove those fake reviews, this Fake Product Review Monitoring system is introduced. The basic idea is to track the IP address of the user submitting the review and if the same kind of review is coming from identical IP address, then those reviews are removed. The system also rates a product based on the reviews submitted by the customers and recommends the best products to the users. The product evaluation is also done, which helps the people to get to know about the product, which they might not have bought earlier. The system incorporates machine learning and opinion mining technology to rate a product.

II. LITERATURE SURVEY

A. An Overview on Research Challenges in Opinion Mining and Sentiment Analysis

P. Kalarani and Dr. S. Selva Brunda [1] mentioned the basics of Opinion Mining and Sentiment Analysis. We also learnt the differences between them. The workflow of Opinion Mining and the classifications of Sentiments are provided. Though these methods look promising and helpful, there are some challenges, when it comes to research or project. The application areas and the tools for Opinion Mining were also discussed. Some major challenges were cited and the possible solutions were also given, which are in developing stage.

B. Toward Social media Opinion Mining for Sustainability Research

Rundong Du et al. [2] showed how the tweets of the people from the popular social networking site Twitter, is used to mine the opinions of the people. The tweets are analysed, whether they are positive or the negative ones. The framework for implementing Opinion Mining in the Twitter to analyse the tweets is given, that helped us to understand the process step-by-step. Some of the real time incidents were discussed as case studies, which were really

challenging for the Natural Language Processing to identify the real meaning for the tweet.

C. Fake Product Review Monitoring and Removal for Genuine Ratings PHP

Manleen Kaur Kohli et al. [3] demonstrated how the present day system permit users to submit their reviews and ensure that all reviews are submitted by a legitimate user. But, the contents of the reviews are not analyzed by the administrators for identifying fake reviews. Various methods have been introduced to identify the fake reviews. One method is by looking at the timestamps and contents of the reviews. Multiple reviews with same timestamp can have similar phrases rearranged in the review content, which imply the same positive/negative feedback regarding a product. These can be identified and marked as fake review.

III. OPINION MINING

[4] Opinion mining is a type of natural language processing for tracking the mood of the public about a particular product.

Opinion mining can be useful in several ways. It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be broadly positive about a digital camera, but be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer.

There are several challenges in opinion mining. The first is that a word that is considered to be positive in one situation may be considered negative in another situation. Take the word "long" for instance. If a customer said a laptop's battery life was long, that would be a positive opinion. If the customer said that the laptop's start-up time was long, however, that would be is a negative opinion. These differences mean that an opinion system trained to gather opinions on one type of product or product feature may not perform very well on another.

A second challenge is that people don't always express opinions the same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In opinion mining, however, "the movie was great" is very different from "the movie was not great".

Finally, people can be contradictory in their statements. Most reviews will have both positive and

negative comments, which is somewhat manageable by analysing sentences one at a time.

IV. PRODUCT RECOMMENDATION SYSTEM

Our system consist of following features they are,

- User will be able to create account.
- User will be able to do online shopping.
- User will be able to write reviews for the purchased products.
- The admin can add or delete the product.
- The admin can remove the fake reviews, by monitoring the IP address.
- The admin can rate the product by the reviews given by the users.
- The admin can recommend the product to the users.

This website is built using web designing languages like HTML, CSS, JavaScript and Bootstrap. The website is completely user friendly. Firstly, users are asked to register to the site if they wish to purchase any product available on the site. Only registered users can buy the products.

Any user who has purchased the product can review it. However, the product manufacturers can take advantage of this fact and can post multiple good reviews of their product to promote it and make it popular. But, the product would be very bad. This results fake reviews. To tackle this situation, our system tracks the IP address of the user submitting the review of the product. Then the administrator will go through the database containing the reviews. The database will contain the reviews, user details i.e., the user who submitted the review along with the tracked IP address. The administrator will search for the instances where multiple reviews implying the same meaning have been submitted by the same IP address. These reviews are identified as fake reviews and the administrator will remove those reviews.

The system rates a product based on the user reviews of the products. [5] This is done using the machine learning and opinion mining techniques. The classification will consist of five classes, representing one for each of 5-star rating i.e., class1 for 1-star rating, class2 for 2-star rating, class3 for 3-star rating, class4 for 4-star rating and class5 for 5-star rating. Firstly, a classifier is built using the 'sklearn' machine learning library. The LinearSVC (a linear support vector machine) classifier imported from sklearn library is used as a classifier. The classifier is trained using a huge database containing the reviews and its ratings. [6] The database used to train the classifier was downloaded online which was available freely which contained thousands of amazon product reviews and its ratings. Then the trained classifier was used to rate the product on a scale of 5 using the user reviews. Initially, to train the classifier and to predict the rating of a review, the review must be represented

numerically and should be given as input to the classifier. This is accomplished using the TfidfVectorizer imported from sklearn feature extraction library. Also, pre-processing of the review text is done by removing the stopwords (which do not convey any sentimental meaning). This task of pre-processing is accomplished using nltk library which contains English stop words. The classifier cannot be trained based on single words in review text, because consider the case when the user review is “not good”. This must be given a 2 or 3-star rating since it is a negative review. But if we train based on single words, this will be given a 4 or 5-star rating after recognizing the word “good” in the review. This case is handled by specifying minimum and maximum range for n-gram attribute in TfidfVectorizer. So, if we specify maximum as 2 for n-gram in TfidfVectorizer, the case where a positive word is negated such as “not good” is handled. This is how a product is rated based on the reviews submitted by the users.

V. ARCHITECTURE OF PRODUCT RECOMMENDATION SYSTEM

The two main components involved in the architecture are,

- User module
- Admin module

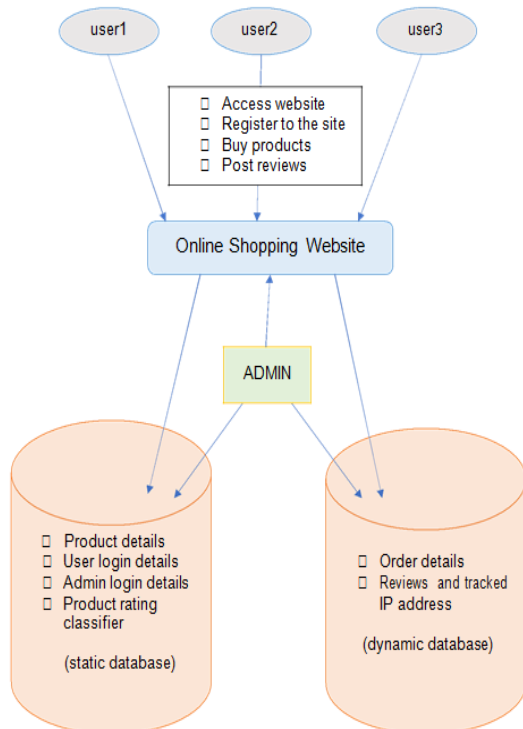


Fig1. Architecture of the proposed system

A. User Module

The users can log in to the website, if they are already registered themselves. If not registered, then they can sign up.

1. Buying the product

The products will be displayed under the respective categories. The users can buy the products, which is displayed along with the image, name and price. The image can be enlarged on hovering to have a clear view of what they are about to buy. Once completing the purchase, they can see the summarized purchase order. After buying, the user has to log out.

2. Reviewing the product

A separate tab has been given in the home page of shopping site for writing reviews. The reviews can be written only by the registered users, only for the products purchased by them. They are asked to login to review window, to ensure the security and authentication. The user can write the review for a product only once and cannot edit the review.

B. Admin Module

The admin is the backbone for this website, as the admin has all the privileges. The admin has the specific login credentials, which cannot be accessed by the others. The admin can add product, delete product, remove fake reviews and rate the products.

1. Adding the product

The products displayed for the users, can be added only by the admin. The products can be added by providing the product id, product name, product description, product image, product price, etc.

2. Deleting the product

The products can also be deleted, only by the admin. The products can be deleted by giving the correct product id or the name of the product. Once the product is deleted, the database gets updated automatically and the deleted product can be added again using Add Product option.

3. Removing Fake review

The fake reviews can be deleted by the admin, if the reviews are found to be fake, meaningless or abusive. The admin has the access to view the reviews, given by the users. If two or more reviews come from a single IP address, then the admin can remove the review.

4. Rating the product

The review for the product is given by the user, but the rating is done only by the admin. When the admin login to his module, he is able to rate the product. The support vector machine algorithm is behind the Rating option. When the admin rates the product, the stars will get assigned according to the percentage of how much the test data matches with the trial data.

VI. CONCLUSION

The Product Recommendation system helps the users to buy the best product that matches their taste. This system also removed the fake reviews by monitoring the IP address, which makes the product genuine and worthy.

VII. FUTURE WORK

In future, this system can be changed into the Personalized Product Recommendation system that can read the user's search history and the purchase behavior, and suggest the products to buy accordingly.

REFERENCES

- [1] P. Kalarani and Dr. S. Selva Brunda, "An Overview on Research Challenges in Opinion Mining and Sentiment Analysis", IJIRCCCE, Vol.3, Issue 10, October 2015
- [2] Rundong Du, Zhongming Lu, Arka Pandit, Da Kuang, John Crittenden, Haesun Park, "Toward Social Media Opinion Mining for Sustainability Research", Papers from the 2015 AAAI Workshop
- [3] Manleen Kaur Kohli, Shaheen Jamil Khan, Tanvi Mirashi, Suraj Gupta, "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews using Opinion Mining", IJARCSSE, Volume 7, Issue 1, January 2017
- [4] <https://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining>
- [5] Product Recommendation System using Opinion mining – IJAERD Vol.4 Issue-1 January 2017
- [6] Monitoring the best product from Consumers opinion forums based on their Budget using Data Mining Technique – IJETT Volume-45 Number-2-March 2017