Detecting Fake Reviews through Opinion Mining

Shaikh Zaitoon[#1], Shaikh Saleha[#2], Qureshi Sohima[#3], Smruti Patil[#4]

[#1-4] IT Department, Padmabhushan Vasantdada Patil Pratishthan's College of Engineering, Sion Mumbai
Department of Information Technology

Abstract: — Recently, Individuals and organizations rely heavily on social media these days for consumer reviews in their decision-making on purchases. However, for personal gains such as profit or fame, people post fake reviews to promote or demote certain target products as well as to deceive the reader. To get genuine user experiences and opinions, there is a need to detect such spam or fake reviews. This paper presents a study that aims to detect truthful, useful reviews and ranks them. An effective supervised learning technique is proposed to detect truthful and useful reviews and rank them, using a 'deceptive' classifier, 'useful' classifier, and a 'ranking' model respectively. Deceptive and non-useful consumer reviews from online review communities such as amazon. comand Epinions.com are used. The proposed method first uses the 'deceptive' classifier to find truthful reviews followed by the 'useful' classifier to find whether a review is useful or not. Manually labeling individual reviews is very difficult and time consuming. We incorporate a dictionary that makes it easy to label reviews. We present the experimental results of our proposed approach using our dictionary with 'deceptive' classifier and 'useful' classifier.

Product reviews play an important role in deciding the sale of a particular product on the e- commerce websites or applications like Flipkart, Amazon, Snapdeal, etc. In this paper, we propose a framework to detect fake product reviews or spam reviews by using Opinion Mining. The Opinion mining is also known as Sentiment Analysis. In sentiment analysis, we try to figure out the opinion of a customer through a piece of text. We first take the review and check if the review is related to the specific product with the help of Decision tree. We use Spam dictionary to identify the spam words in the reviews. In Text Mining we apply several algorithms and on the basis of these algorithms we get the specific results.

1. Introduction

Opinion Mining (OM), is the domain of study that analyzes people's opinions, evaluations, sentiments, attitudes, appraisals, and emotions towards entities such as services, individuals, issues, topics, and their attributes [1]. "The sentiment is usually formulated as a two-class classification problem, positive and negative". Sometimes, time is more precious than money, therefore instead of spending time in reading and figuring out the positivity or negativity of a review, we can use automated techniques for Sentiment Analysis.

Nowadays, consumers looking to buy a product increasingly rely on user-generated online reviews to make or reverse their purchase decisions. Positive reviews of a product greatly influence the person's decision to buy the product. However, if one sees many negative reviews, he/she will most likely choose a different product. The outcome of positive reviews gives significant profit and advertizing for the seller and their organization. This in turn creates a market for incentivizing opinion spam. This has resulted in more and more people trying to game the system by writing fake reviews to harm or promote some products or services. A fake review means that it iseither a positive review written by the business owners themselves (or people they contract to write reviews) or a negative review written by a business's competitors. Those fake reviews try to deliberately mislead readers by giving fake reviews to some entities (e.g. products) in order to promote them or to damage their reputation.

The main contributions of this study are summarized as follows:

- Using the Weka tool [29], we compare different sentiment classification algorithms which are used to classify the movie reviews dataset into fake and real reviews.
- We apply the algorithms using two different datasets with stopwords. We realized that using the stopwords method is more efficient than without stopwords not only in text categorization, but also to detection of fake reviews.
- We perform several analysis and tests to find the learning algorithm in terms of accuracy.

2. Literature Survey

The purpose of a literature review is to gain an understanding of the existing research and debates relevant to a particular topic or area of study and to present that knowledge in the form of a written report.

Conducting a literature review helps you build your knowledge in your field.

A. Data Acquisition:

In this step, we prepare an in-house data set of spam reviews and reviewers using human collected from online e-commerce websites or application like Amazon, Flipkart with different characteristics and sizes. The records are chosen randomly from any of the records that are available on the website.

B. Data Integration:

In this step, we combine the data from multiple review source data sets into a coherent form.

C. Spam Identification Labelling:

In this step, we look for various types of the spam in the data integrated set, and labeled each record as spam and non-spam manually.

D. Pre-processing:

In this step, we use various types of pre-processing techniques to handle the missing, noisy and inconsistent data. There are a number of pre-processing techniques such as case folding dam character erase, tokenization, slang word handling, stop word removal, stemming and number handling.

E. Evaluate the Approach:

In this step, analyzation of the outcome and rationalize the feasibility of the approach we followed by comparing it with other previous approaches. The Public Opinion Survey conducted consisted of several questions relating to the purpose about analyzing the importance of product review for various customers while buying any product. The participants in the survey are mostly the frequent online buyers. The survey is designed as such to know the public opinion about referring to the product review while buying it. The result of the survey was as expected and was in favors of the arguments proposed earlier.

3. Existing System

By conducting several public opinion surveys, based on their results it can be evaluated that people do read and get influenced by ratings and reviews of the products online. A survey performed by a leading site has shown that: More than 80% of the online customers look at the reviews available. 50% base their purchase on the ratings of the products. 30% of the customers compare the ratings of similar products before making their Decision. Clearly consumers value the feedback given by other users as do the companies that sell such products. Blogs, websites, discussion boards etc. are a repository of customer suggestions which are valuable and important sources of textual data. Therefore, today's individuals and older ones extensively rely on reviews available online. It means that people make their decisions of whether to purchase the products or not by analyzing and reflecting the existing opinions on those products. The fact that is if the potential customer or users gets a genuine overall impression of a product by considering the present affect for that product, it is highly probable that he will actually purchase the product. Normally if the percentage of positive and effective opinions is considerable, it is likely that the overall impression will be highly positive. Likewise, if the overall impression is not proper, it is doubtful that they don't buy the Product. Now the customers can write any opinion text, this can motivates the individuals, and organizations to give undeserving spam opinions to promote or not to credit some target products.

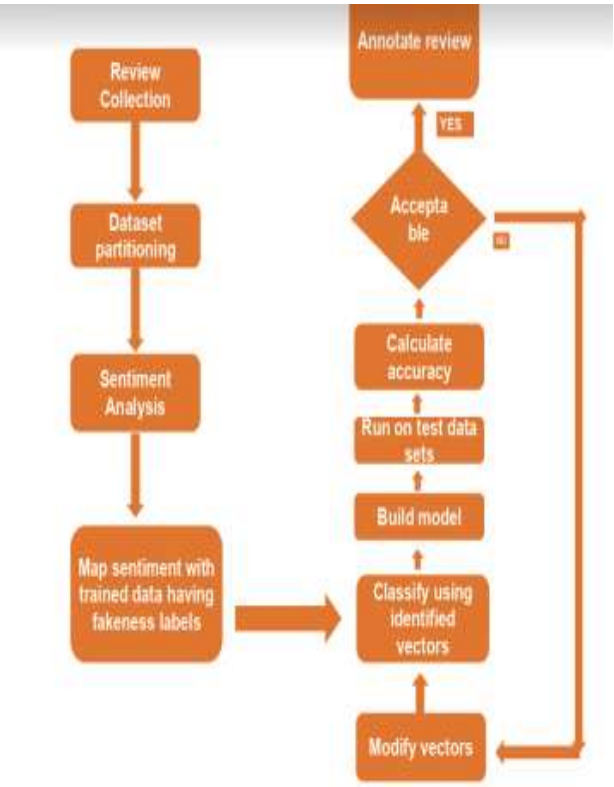

Figure 1.4 Work flow

4. Proposed System

The main agenda is to further improve customer satisfaction and online shopping experience. In order to do so, it has become a common practice for online merchants to enable their customers to put forward their reviews on the products that they have purchased. With more computer users becoming comfortable with the Web, a huge number of people are coming forward to write the reviews and post them on website which is becoming beneficial for other customers. It also decides profit or loss for any e-commerce merchant. As an outcome of this, the number of reviews that a product receives is growing quickly. Most of the famous products get thousands of reviews at some large merchant sites. Now any customer can write any opinion text or review, this can draw the individual's attention and organizations to give undeserving spam opinions to promote or to discredit some target products. The existing system doesn't restrict spam and invalid reviews and comments. So there is a need to develop a smart system which automatically mine opinions and classify them into spam and non-spam categories.

5. Conclusion

Opinions are very important for anyone who is going to make a decision. Web mining has emerged in recent years as an attractive technology to individuals and corporations to know others opinions. Opinion mining is helpful for individuals when they want to buy a product and they can decide which product to buy, by studying the summarized opinions instead of studying long reviews and making summary themselves. Review text is an important source of information for the consumers before purchasing any product from e-Commerce websites. In recent years, review spam has received significant attention in both business and academia. In this project, we investigate the need from the blind and visually impaired people. Base on the impetus of CNN, we develop a blind visualization system that helps blind people better explore the surrounding environment. A portable and real-time solution is provided in the project. We present a platform that helps to transform the visual world into the audio world with the potential to inform blind people objects as well as their spatial locations. We introduce YOLO, a unified model for object detection. Our model is simple to construct and can be trained directly on full images. Unlike classifier-based approaches, YOLO is trained on a loss function that directly corresponds to detection performance and the entire model is trained jointly.

6. Acknowledgement

The physical support and the technical contribution of Prof. Smruti Patil is highly appreciable without their moral support it was impossible to reach the goal.

7. References

[1] B. Liu, "Sentiment analysis and opinion mining," Synthesis lectures onhuman language technologies, vol. 5, no. 1, 2012, pp. 1–167.

[2] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithmsand applications: A survey," Ain Shams Engineering Journal, vol. 5, no. 4, 2014, pp. 1093–1113.

[3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classi-fication using machine learning techniques," in Proceedings of EMNLP,2002, pp. 79–86. [Online]. Available

:http://www.cs.cornell.edu/People/pabo/movie%2Dreview%2Ddat a/