Spam Detection using Sentiment Analysis

Aishwarya Ashok Aagte
Department of Computer Engineering
Vidyalankar Institute of Technology
Mumbai,India

Vrushali Wagh
Department of Computer Engineering
Vidyalankar Institute of Technology
Mumbai,India

Prince Vishwakarma
Department of Computer Engineering
Vidyalankar Institute of Technology
Mumbai,India

Prof. Sandeep Kamble
Department of Computer Engineering
Vidyalankar Institute of Technology
Mumbai,India

Abstract— **the rapid growth of internet has made us dependent on it for most of our activities. One such growing area is e – commerce. Today, there are hundreds of e – commerce sites and millions of products available on them. These sites allow its users to write a review about the product or the service provided. These reviews can act as a great source of information for the service provider for analyzing sales or decision making. It can also be useful for potential users to decide whether to buy a product or not. However, these reviews can be useful only if they are true. Rapid growth of e – commerce has also led to great competition among companies/brands. Hence, some parties try to add fake reviews on the competitor's site in order to raise the popularity or to criticize a product/brand. This paper talks about the various techniques for detecting spam.**

Keywords—sentiment analysis, opinion mining, Naïve Bayes, spam review

I. Introduction

In the recent past years, the use of online shopping has increased tremendously. Most retail website provide platform to post reviews on millions of product and encourage people to post their opinion or sentiment on various aspect of the product. The reviews play very important role while purchasing product, because consumer can make wise purchasing decision by paying more attention towards important aspect and companies will concentrate on important aspect while improving the quality. Let's see a sample review "The battery life of blackberry curve is amazing." "This camera has excellent picture quality." "The battery life of micromax a94 is not good." The above reviews contain opinion about battery and camera. The first review about battery is a positive opinion and the last review about battery is a negative opinion. About the camera there is a positive opinion.

There are millions of reviews on millions of product that are available on various websites. Generally product may have numbers of aspect. For example mobile has aspect like usability, design, applications, network, battery etc. for laptop, the aspect such as hard disk, RAM, Graphics card, screen, Battery etc. The product aspects are greatly influenced on product quality. So aspect identification is very important task while buying product. Paying more attention to the important aspect is very useful while taking decisions about product. Companies or brands can focus on improving and enhancing the quality of product aspect and enhance the reputation of the product more effectively. It is very difficult for customer to identify the important aspect of product from various websites. And also the reviews are often disorganized it causes problem while knowledge acquisition about product. In our proposed work, we have attempted to detect spam and fake reviews by filtering out vulgar, expletives and curse words by using sentiment analysis.

II. Literature Survey

A. *Types of spam -*

A number of studies have been conducted which focused on spam detection in e-mail and on the web, however, only recently have any studies been conducted on opinion spam. Jindal and Liu (2008) have worked on "Opinion Spam and Analysis" [1] and have found that opinion spam is widespread and different in nature from Web spam. They have classified spam reviews into 3 types: Type 1, Type 2 and Type 3.

Type 1 spam reviews are untruthful opinions that try to mislead readers or opinion mining systems by giving untruthful reviews to some target objects for their own gains.

Type 2 spam reviews are brand only reviews, those that comment only on the brand and not on the products.

Type 3 spam reviews are not actually reviews; they are mainly either advertisements or irrelevant reviews which do not contain any opinions about the target object or brand. Although humans detect this kind of opinion spam they need to be filtered, as it is a nuisance for the end use

B. *How different websites deal with fake/spam reviews -*

i. Amazon -

Amazon does remove fake reviews, but does not disclose much about the methodology. This is commonly done for content moderation, because the more one discloses about the methods, the easier it becomes for unscrupulous types to game the system.

One way that reviews reaches Amazon is through customer feedback. Near every review, there's a link to report abuse. This feature can be used to report reviews which violate guidelines, including fake reviews. Amazon very rarely removes reviews and when they do, it's because the customer have directly, explicitly, and undoubtedly violated their terms.

For example, if there is a curse word in the review, it'll likely be removed. However they are only able to identify a portion of them as it is not so easy to distinguish between fake and genuine reviews. For example, if a buyer creates username and gives two positive reviews for a single company and never review again, it is likely that reviews are fake, but there is some small chance that the customer is a real person who just only used Amazon one time and they would not want to upset that customer by deleting my profile and reviews.

Amazon only reacts to complaints and without any verification process removes the review. If someone enquires on it Amazon will send a canned response.

ii Flipkart

Flipkart does not influence ratings and reviews on the platform. The Flipkart team creates a portfolio for each product before it goes live on their page. Every possible detail right from the product images, specifications, to description and other details are updated on the site.

Flipkart reviewers can grade a product on a scale of 1(being the lowest) to 5 stars (being the highest). The products popularity is higher among shoppers when the reviews and the average rating of that product is high.

This is how Flipkart treats reviews -

- **Top reviews:** Under each review submitted by a shopper is an option for readers to judge its utility value. Reviews with a positive response to the question 'Was this review helpful?' are ranked higher as they are more likely to be authentic. This logic is not influenced by Flipkart but stems from buyer or visitor feedback.
- **Star rating:** Users who write reviews on Flipkart can also rank products on a scale of 1 to 5 stars. Rating 1 is for poor review while rating % denotes excellent product review. One should look for both low and high star-rated reviews for a better idea of product quality.
- **Certified buyer reviews:** You might have observed that certain product reviews on Flipkart are identified by a dark green horizontal band with the word 'Certified' emblazoned on it. These banners ensure that the reviews are authentic and also identifies that the review is from a certified buyer. Only certified buyers are allowed to write product reviews on platform.

In order to address the issue of fake/spam reviews Flipkart has a dedicated team that examines certain parameters to identify fake reviews and rating. It follows the below guidelines:-

- Detecting sellers that pose as buyers on the platform and boost ratings and reviews of their own products
- Identifying competing sellers posing as buyers and posting fraudulent reviews on the rival's product page
- Sellers that pose as buyers and boost their own seller ratings and reviews
- Sellers that pose as buyers and try to pull down their competition by posting negative seller ratings and negative reviews
- Third party vendors hired by sellers to write reviews that boost their product ratings and reviews.

If any of the above fraudulent activity is found, then Flipkart removes such account.

iii Zomato

Zomato has worked relentlessly to identify both restaurants and users who engage in unethical means to influence restaurant ratings and reviews. In recent years, Zomato's neutrality team has removed over 300+ high activity users from the Zomato platform on grounds of solicitation where influential foodies or food bloggers are incentivized or offered to join hands with restaurants to write good things about restaurants. In return, these reviewers get monetary gains. The team of Zomato has found enough evidences of unethical practices to remove such accounts.

Zomato has now shifted its focus to restaurant owners who also play an active role in promoting solicitation. They have started warning users of suspicious reviews and ratings for such restaurants on the Zomato platform. This will done through a two step process –

1. Zomato will be sending a warning to restaurants where they will detect suspicious activity – restaurants will have to respond quickly to make sure that all solicited reviews are proactively detected and removed from the Zomato platform and their listing page.
2. Post such warnings, if any malicious behavior continues, Zomato will display this warning banner on the restaurant's page on Zomato app. For a period of three months this banner will remain on the page.

These are some steps taken by Zomato in order to safeguard their platform.

III. Sentiment Analysis

Sentiment analysis uses natural language processing (NLP), text analysis to identify, analyze, quantify subjective information. It's application ranges from customer service to marketing which includes reviewing survey responses and reviews, online and social media, etc.

It is also known as Opinion Mining which is the domain of study that analyzes people's opinions, evaluations, sentiments, attitudes, appraisals, and emotions. [5]

The basic of Sentiment Analysis is classifying the polarity i.e. reviewing whether the given text is positive negative or neutral. The given text is in 3 levels – document, sentence and aspect.

The aim of sentiment analysis is to find opinions from reviews and classify these opinions based on polarity.

Opinions are classified into three categories [2]:

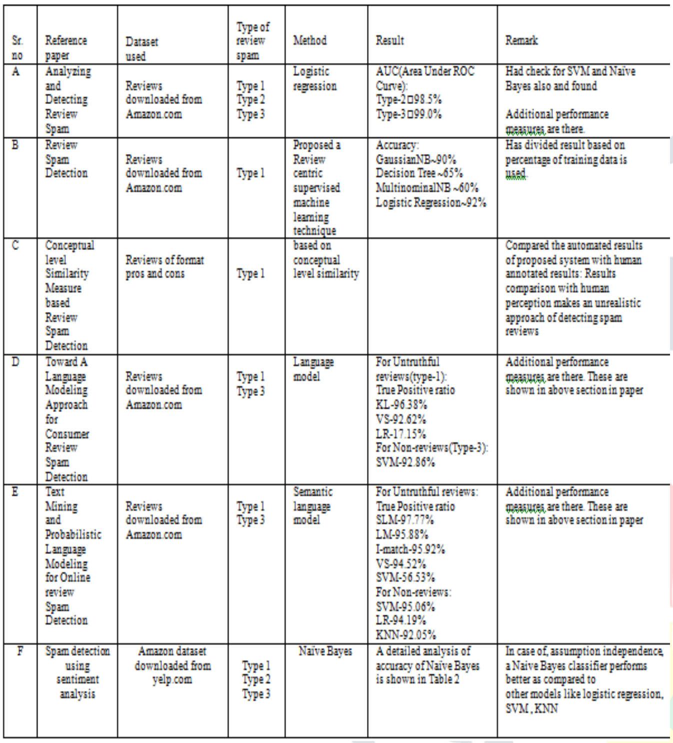
i. Direct opinions which opinion holder directly attack to target.

ii. Comparative opinions which are opinion holder compare among entity.

iii. Indirect opinions, which are implied as in idioms or expressed in a reverse way as in sarcasm.

IV. ANALYSIS OF SPAM DETECTION TECHNIQUES

The comparison and analysis of all the techniques used in previous paper is shown as [3]



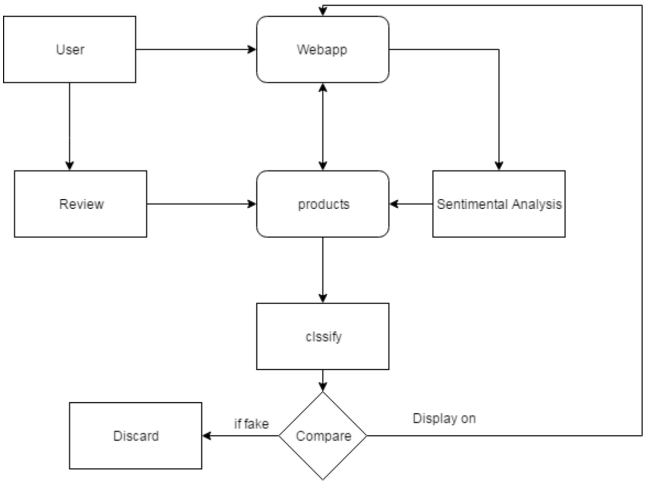
Sr. no	Reference paper	Dataset used	Type of review spam	Method	Result	Remark
A	Analyzing and Detecting Review Spam	Reviews downloaded from Amazon.com	Type 1 Type 2 Type 3	Logistic regression	AUC(Area Under ROC Curve): Type-2□98.5% Type-3□99.0%	Had check for SVM and Naïve Bayes also and found Additional performance measures are there.
B	Review Spam Detection	Reviews downloaded from Amazon.com	Type 1	Proposed a Review centric supervised machine learning technique	Accuracy: GaussianNB~90% Decision Tree ~65% MultinominalNB ~60% Logistic Regression~92%	Has divided result based on percentage of training data is used.
C	Conceptual level Similarity Measure based Review Spam Detection	Reviews of format pros and cons	Type 1	based on conceptual level similarity		Compared the automated results of proposed system with human annotated results: Results comparison with human perception makes an unrealistic approach of detecting spam reviews
D	Toward A Language Modeling Approach for Consumer Review Spam Detection	Reviews downloaded from Amazon.com	Type 1 Type 3	Language model	For Untruthful reviews(type-1): True Positive ratio KL-96.38% VS-92.62% LR-17.15% For Non-reviews(Type-3): SVM-92.86%	Additional performance measures are there. These are shown in above section in paper
E	Text Mining and Probabilistic Language Modeling for Online review Spam Detection	Reviews downloaded from Amazon.com	Type 1 Type 3	Semantic language model	For Untruthful reviews: True Positive ratio SLM-97.77% LM-95.88% I-match-95.92% VS-94.52% SVM-56.53% For Non-reviews: SVM-95.06% LR-94.19% KNN-92.05%	Additional performance measures are there. These are shown in above section in paper
F	Spam detection using sentiment analysis	Amazon dataset downloaded from yelp.com	Type 1 Type 2 Type 3	Naïve Bayes	A detailed analysis of accuracy of Naïve Bayes is shown in Table 2	In case of, assumption independence, a Naive Bayes classifier performs better as compared to other models like logistic regression, SVM , KNN

Table 1. Analysis of spam detection techniques

V. PROPOSED METHODOLOGY

The goal is to incorporate sentiment analysis into spam review detection. To achieve this, first the user will review a product and view the reviews that are sorted by the admin. After the user comments on a product, the admin will sort the reviews and remove those reviews that are spurious.

.



- Following steps are carried out -

1) Reviews extraction and Preprocessing.

2) Aspect Identification of the product.

3) Using sentiment classifier for classifying the positive and negative reviews of product.

4) The Naive Bayes algorithm is used for ranking

Preprocessing the review
↓
Performing Tokenization
↓
Removing stop word
↓
Performing Lemmatization
↓
Passing it through classifier
↓
Getting labels

1. *Reviews Extraction and Pre-processing:*

The first step is data preprocessing which is very important task to be done before product aspect identification. Reviews are generally less formal and written in an ad hoc manner as compared to regular text document. If sentiment analysis is applied on raw review then often the performance achieved is very poor. Therefore the preprocessing techniques on reviews are necessary for obtaining satisfactory result on sentiment analysis.

2. *Aspect Identification of the product:*

In Aspect identification we identify aspect from numerous customer reviews. The reviews are available on different forum websites. But customer reviews are composed in different formats on various forum websites. Customer review consists of positive, negative or neutral reviews. On some website the reviews are in free text paragraph format and on some website there is an overall rating on the product. The aspects of the product are identified as a frequent Noun term from these reviews.

3. *Sentiment Classifier:*

The aim of sentiment classification is to classify the given text to one or more predefined sentiment categories. The categories can be Positive, Negative, Neutral. The NPL techniques are used to find out the customer reviews from their own languages and it is also for converting it into a format that's understandable.

4. *Aspect Ranking Algorithm*

This will identify the important aspect of product from online customer reviews. The important aspects are commented again and again in the review and the customer's opinions on the important aspect greatly influence their overall opinions on the product. The aggregation of the opinions given to specific aspects in the review is an overall opinion in a review. Various aspectsc contribute differently in the aggregation. That is, the opinions on important aspects have strong impacts on the generation of overall opinion and vice versa.

Different approaches to classify the machine learning methods which includes Naïve Bayes or Support Vector Machine.

5. *Naïve Bayes*

It is very simple and useful classifier based on Bayes theorem of probability. Naïve Bayes (NB) classifier is a basic probabilistic classifier with strong independent assumptions. It calculates a set of probabilities by combinations of values in a given dataset.

NB is very simple and efficient and technique for spam filtering. Creating a Naïve Bayes network helps in exploiting commonness among different tasks, thus learning and modifying accordingly. [9]

It is applicable for document level classification which uses independence between the objectives. Naives Bayes improved version solves problem like tendency, where correctness of positive word appears approximately 10% more accurate than negative word. [4]

$$P(S|W) = \frac{P(W|S)\,P(S)}{P(W|S)\,P(S) + P(W|NS)\,P(S)}$$

Where,

W – word in a review
S – Spam review
NS – non spam review

- Some advantages of Naïve Bayes are :-
  - In case of, assumption independence, a Naive Bayes classifier performs better as compared to other models like logistic regression
  - There are fewer requirements of training data.
  - It is not most interoperable but more interoperable than k-nearest neighbors.
  - It handles missing data very well.
  - Also, the NB classifier has fast decisions making process.
  - Since this classifier returns probabilities, it is simpler to apply these results to a wide variety of tasks than if an arbitrary scale was used [8]
  - It is very intuitive. Unlike neural networks, they do not have several free parameters that must be set. [8]

Sr no.	Paper	Publication & Year	Analysis
1.	Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool.	IJACAS 2013	Naïve Bayes is the most accurate compared to Decision Tree and k-NN with the average accuracy of 0.737. The average accuracies of Decision Tree and k-NN are 0.589 and 0.567, respectively. [8]
2.	Spam Filtering using Hybrid local – global Naïve Bayes Classifier.	IEEE 2015	This paper observed high accuracy rate with small deviation in classification of messages. Nearly 95% of spam and 93% of ham messages were correctly classified. [9]
3.	Detecting Fake Reviews utilizing Semantic and Emotion Mode.	IEEE 2016	This paper found that when review density, semantic, emotional features were used Naives Bayes gave a precision of 0.81 while SVM and Decision gave a precision of 0.92 and 0.93 respectively. [6]
4.	An efficient Naïve Bayes with Negation Handling for Seismic Hazard Prediction.	IEEE 2016	As per this paper, NB with negation handling (76.98%) performed well as compared o MATLAB Native NB (65.09%). [10]
5.	The Impact of training dataset on the Accuracy of Sentiment Classification of Naïve Bayes Technique	IEEE 2017	Based on 5 experimental dataset of 5,10,25,50,100 tweets dataset this paper concluded that, the accuracy level of analysis increases as the no. of dataset of training data increases up to a certain point. The average accuracy of these 5 experiments is about 76% with a deviation of 0.1360. [11]
6.	Comparison of Naïve Bayes and Support Vector Machine Classifiers on document classification.	IEEE 2018	According to this paper, NB gave 81% and SVM gave 85% accuracy results. [12]

Table 2. Analysis of Naïve Bayes accuracy

VI. Conclusion

This paper proposes incorporation of sentiment analysis to detect spam reviews. It also talks about the how different websites and mobile app analyze and treat spam/fake review. This paper also includes details about pre-processing data before aspect identification. Aspect identification classifies the review as positive, negative or neutral. Finally, Naïve Bayes is used for aspect ranking.

VII. Future Scope

In future we can try to develop methods to calculate sentiment score that are more efficient. Also, the dictionary containing the sentiment words can be updated. Developing system where there is computer assisted labeling of reviews so as to reduce the workload of humans.

Acknowledgment

For publishing this paper, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. We would like to show our gratitude to Dr. Arun Chavan, Head of Computer Department, V.I.T. for his useful insights. We would also like to thank Prof. Sandeep Kamble for giving us a good guideline throughout numerous consultations.

References

[1] Jindal Nitin and Liu Bing, “Opinion spam and analysis” , Proceedings of the 2008 International Conference on WebSearch and Data Mining, New York, ACM press, 2008:219-230.

[2] Suad Alhojely “ Sentiment Analysis and Opinion Mining: A Survey”, proceedings of International Journal of Computer Applications (0975 – 8887) Volume 150 – No.6, September 2016.

[3] Prof. Ankit P. Vaishnav and Vyas Krishna Maheshchandra “A survey on Review Spam Detection Techniques”, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 4 Issue 04, April-2015.

[4] Chirag Visani and Navjyotsinh Jadeja “A study ofn Different Machine Learning techniques for spam review detection”, Conference Paper · August 2017.

[5] Elshrif Elmurngi and Abdelouahed Gherbi “An Empirical Study on Detecting Fake Reviews Using Machine Learning Techniques” , The Seventh International Conference on Innovative Computing Technology (INTECH 2017)

[6] Yuejun Li, Xiao Feng and Shuwu Zhang “Detecting Fake Reviews utilizing Semantic and Emotion Mode”, 2016 3rd International Conference on Information Science and Control Engineering.

[7] Shashank Kumar Chauhan, Anupam Goel, Prafull Goel, Avishkar Chauhan and Mahendra Gurve “Research on Product Review Analysis and Spam Revie Detection”, 2017 4th International Conference on signal Processing and Integrated Networks (SPIN).

[8] Ahmad Ansari, Iman Paryudi, A Min Tjoa “Performance Comparioson between Naïve Bayes, Decision Tree and k-Nearest Neigbor in Searching Alternative Design in an Energy Simulation Tool” (IJACSA) International Journal of Advanced Com[puter Science and Applications Vol.4, No. 11, 2013

[9] Rohit Kumar Solanki, Karun Verma and Ravinder Kumar “Spam Filtering using Hybrid local – global Naïve Bayes Classifier” IEEE 2015.

[10] Klyan Netti and Dr. Y Radhika “An efficient Naïve Bayes with Negation Handling for Seismic Hazard Prediction” IEEE 2016.

[11] Mohd Naim Mohd Ibrahim and Mohd Zaliman Mohd Yusoff “The Impact of training dataset on the Accuracy of Sentiment Classification of Naïve Bayes Technique” 2017 IEEE Conference on Open Systems (ICOS).

[12] ZUN Hlaing Moe, Thida San, Mie Mie Khin and Hlaing May Tin “Comparison of Naïve Bayes and Support Vector Machine Classifiers on document classification” 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE 2018)