ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Detection of Spam Reviews in Mobiles

Saiesh N. Prabhu Verlekar

Assistant Professor, Dept. of I.T., SRIEIT, Goa University, Goa, India

ABSTRACT: It is now a common practice for e-commerce web sites to enable their customers to write reviews of products that they have purchased, reviews provide valuable sources of information on these products. Reviews are analyzed by customers who are interested in buying new mobiles and also by manufacturers to find the faults in their product. This importance of reviews also gives good incentive for spam, which contains false positive or malicious negative opinions. Review spam is quite different from web page spam and email spam, and thus requires different detection techniques. The threats posed by spam review are that the manufacturer may conclude about their product incorrectly and potential buyers will be reduced. The objectives of this paper is to allow the customers to view legitimate reviews thus influencing their decision to purchase good products thereby trusting the reviews that are provided by other customers.

KEYWORDS: Review spam, Product Reviews.

I. INTRODUCTION

With the development of the Internet, people are more likely to express their views about products on the web. They can now post reviews of products at e commerce sites and express their views in blogs and forums. There was a growing interest in mining opinions from reviews. However, the existing work is mainly on extracting positive and negative opinions using natural language processing techniques. There is no reported study on the trustworthiness of reviews, which is crucial for all opinion based applications. There is no quality control and anyone can write anything on the web, which results in many low quality reviews, and worse still spam reviews. Recognizing whether a review is a spam review or not is extremely difficult by manually reading the reviews because one can carefully craft a spam review which is just like any other innocent review and the number of spam reviews is also small. The spam means use of electronic messaging system to send unsolicited, irrelevant bulk messages. Review spam is one such spam that occurs due to false positive or malicious negative reviews. There are two types of spam reviews Misleading reviews and Non reviews. Misleading are the ones that give undeserving opinion that deliberately misguide the people. Non reviews are the ones which contain no opinion but some advertisements or chat having no connection with the product. As the spam reviews can cause people to conclude incorrectly it becomes necessary to eliminate the spam. Once the spam reviews are identified they can then be filtered leaving behind the non-spam reviews.

II. LITERATURE REVIEW

Web Scraping

Web scraping is the process of automatically collecting web information. Web scraping focuses more on the transformation of unstructured data on the Web, typically in HTML format, into structured data that can be stored and analysed in a central local database or spread sheet. Web scraping is also related to Web automation, which simulates human web browsing using computer software.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 6, June 2017

Preprocessing

Preprocessing of Review Spam Detection involves three phases:

- Tokenization
- Stopword Removal
- Stemming

Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing

All contiguous strings of alphabetic characters are part of one token; likewise with numbers. Tokens are separated by whitespace characters, such as a space or line break, or by punctuation characters. Punctuation and whitespace may or may not be included in the resulting list of tokens.

Stopwords are common words that carry less important meaning than keywords. Usually search engines remove stopwords from a keyword phrase to return the most relevant result. i.e. stopwords drive much less traffic than keywords.

Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root for generally a written word form. The stem need not be identical to the morphological root of the word. For stemming, porter stemmer algorithm is used.

Naive Bayesian Algorithm

The Naive Bayes algorithm is based on conditional probabilities. It uses Bayes Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data.

Bayes Theorem finds the probability of an event occurring given the probability of another event that has already occurred. If B represents the dependent event and A represents the prior event, Bayes theorem can be stated as follows.

Bayes Theorem: Prob(B given A) = Prob(A and B)/Prob(A)

To calculate the probability of B given A, the algorithm counts the number of cases where A and B occur together and divides it by the number of cases where A occurs alone.

K-Nearest Neighbour algorithm

K-nearest neighbor algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. *k*-NN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification.

The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms: an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of its nearest neighbour.

The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data.

Advantages of KNN:

- Robust with regard to the search space; for instance, classes don't have to be linearly separable.
- Few parameters to tune: distance metric and k.

Disadvantages of KNN:

- Need to determine value of parameter K (number of nearest neighbors).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

- Distance based learning is not clear which type of distance to use and which attribute to use to produce the best results.
- Computation cost is quite high because we need to compute distance of each query instance to all training samples.

III. PROPOSED METHODOLOGY AND DISCUSSION

Spam reviews are classified into 3 different categories [1] namely Type 1, Type 2 and Type 3. Type 1: Type 1 refers to false positive or malicious negative reviews. False positive reviews are posted by manufacturers to promote their product. Malicious negative reviews damage the reputation of a firm [1]. Type 2: Type 2 reviews consist of the reviews that refer to brand and not the actual product [1]. Type 3: Type 3 review refers to advertisements and other non reviews. The advertisements are posted to deviate the user from actual product and other non-reviews contain question, answer, comments and random text [1].

Review Spam Detection system is divided into 4 modules:

1. Review Extraction
2. Review Pre-processing
3. Review Training
4. Review Classification.

Review Extraction: This module deals with web scraping wherein a scraper performs the operation of matching the required review pattern through the source code of the file. This module basically performs 2 functions:

- Review Region Extraction
- Review Extraction

Review Region Extraction: Extracting the review region is done by identifying the block of HTML code where the reviews are placed by scanning through the entire source code. The irrelevant source code is removed at this step.

Review Extraction: Once the region is identified the pattern for each review is considered and accordingly each review attribute is stored in the corresponding text file which is created dynamically based on number of reviews. The attributes are the name of the reviewer, the content, the date and time of post along with the rating.

Review Pre-processing: Once the reviews are extracted with the help of scraper. The reviews are stored in a corpus. Tokenization and stopword removal techniques are applied on the review. The output what we get is the pre-process data which will be the input to Bayesian filter.

Review Training: This phase involves training the system in such a way that when new review appears during testing phase, the system must be intelligent enough to assign label to that review. As a first step for training, we have classified some of the reviews from the merchant site into spam and ham data sets manually. This manual training is termed as Annotation Training.

Review Classification: Probability of each token is removed based on spam and ham probability formula. If the probability of token is greater than threshold (0.9) it is marked as spam else non spam (ham).Spamicity is calculated using spamicity formula. The formula used to calculate a token's spamicity from these pieces of information is as follows:

$$\text{Ham Probability} = \frac{\text{Token Frequency in ham messages}}{\text{Number of ham messages trained on}} \qquad \text{(eqn.1)}$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 6, June 2017

$$\text{Spam Probability} = \frac{\text{Token Frequency in spam messages}}{\text{Number of spam messages trained on}} \quad \text{(eqn.2)}$$

$$\text{Spamicity} = \frac{\text{Ham Probability}}{\text{(Ham Probability + Spam Probability)}} \quad \text{(eqn.3)}$$

If either Ham probability or Spam probability are greater than 1.0, set them equal to 1.0.

$$\text{Total probability } P = \frac{p1.\ p2 \ldots.pN}{p1.\ p2\ldots.pN + (1- p1)\ (1- p2)\ldots.\ (1-pN)} \quad \text{(eqn.4)}$$

Where: P is the probability that the suspect message is spam.
p1 is the probability p(S | W1) that it is a spam knowing it contains a first word.
p2 is the probability p(S | W2) that it is a spam knowing it contains a second word.
pN is the probability p(S | WN) that it is a spam knowing it contains an Nth word .

The spam review is identified with the help of Naïve Bayesian classification algorithm. Here the test data is considered as an input without a class label and based on the training set label is assigned and those with spam as a label are filtered.

IV. PROPOSED ALGORITHM

Bayesian SPAM filters [3] consider the historical probability of each word in the message occurring in either SPAM [3] or non-SPAM (HAM) [3] messages. They calculate the probability that review is SPAM or non-SPAM by combine the individual SPAM/HAM probability of each word inside the messages to produce a final probability estimate that a review is SPAM or HAM (non-SPAM). The percentage of false positive generated by Bayesian filters are low, and they are self-adapting to stop new SPAM by receiving ongoing training from the user [2].

How it works [2]:

Count up occurrences of words in previous data.
How many times does X appear in spam data?
How many times does X appear in non-spam data (i.e. is ham)?
From these counts, we calculate the spam probability of a word.

For each calculation, you need to know:

P (word|spam) = (# spam messages with word) / (# spam messages)
P (spam) = (# spam messages) / (# total messages)
P (word) = (# messages with word) / (# total messages)
To calculate P (spam|word), Use Bayes' Rule: P (spam|word) =P (word|spam)*P (spam)/ P (word)
To calculate the probability of ham, substitute ham for spam in the equations above [2].

V. RESULTS

This paper was proposed to detect review spam using four modules. Experiment using naïve Bayesian algorithm showed good results. K-NN algorithm was also used on review spam and found less accurate compared to naïve Bayesian.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Website: www.ijircce.com
Vol. 5, Issue 6, June 2017

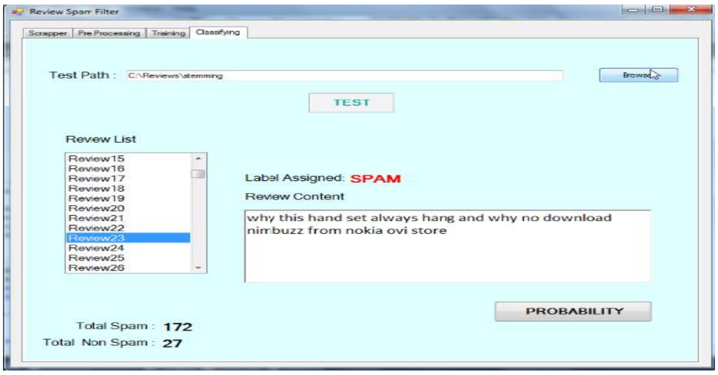

Figure 5.1 Classification using Naive Bayesian

After performing testing on total number of 199 reviews using Naive Bayesian Classifier, the naïve Bayesian classifier produces a total spam of 172 reviews and non spam of 27 reviews.

Figure 5.2 Classification using K-NN

After performing testing on total number of 201 reviews using K-NN, the K-NN classifier produces a total spam of 170 reviews and non-spam of 31 reviews.

VI. CONCLUSION

The Literature survey was conducted on KNN Classifier. Various other rule based and content based algorithms were analysed. It was found that after applying equal numbers of reviews through testing phase KNN Classifier takes lot of time in processing the reviews and only few reviews are labelled as spam compared to Naive Bayesian Algorithm resulting in providing less accuracy compared to Naive Bayesian.
The study of Naïve Bayesian is carried out which is mostly used for spam detection resulting in selection of Naïve Bayesian method as it provided 94% accuracy than any other classification technique.

VII. FUTURE WORK

More and more reviews can be added to training set in order to increase the efficiency of the system. Naive Bayesian Classier can be improved by using some new techniques.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

REFERENCES

[1] Analyzing and Detecting Spam Review - By Bing Liu and Nitin Jindal.
[2] Language and Computers (Ling 384) Topic 3: Language Identification and Spam Detection by Adriane Boyd_Winter 2006.
[3] Taxonomy of email spam filters – Hasan Shojaa Alkahtani, Paul Gardner-stephen and Robert Goodwin.
[4] An Overview of Content-Based Spam Filtering Techniques by Ahmed Khorsi,Department of Computer Science,Djillali Liabes University, Bel Abbes, 22000, Algeria.
[5] Review Spam Detection by Bing Lui and Nitin Jindal.