BOGUS REVIEW ANALYSIS

[1]Kowshika.S, [2]Manohari.R, [3]Naveena.C, [4]Manivannan.D,

Department of Information Technology, Dr.N.G.P. Institute of Technology

[5]Mani Deepak Choudhry, Assistant Professor, Information Technology, Dr.N.G.P. Institute of Technology

[1]kowshisenthilkumar@gmail.com, [2]manoharirajendran@gmail.com, [3]cnaveenayalu@gmail.com, [4]manivannan9198@gmail.com, [5]snsdeepak@gmail.com.

ABSTRACT

While making any purchase online consumer usually checks opinions of others about the product. Manufacturer can gain insight about their products strength and weaknesses based on the reviews of the customers. As people are spending more time to shop and view reviews online, some reviewer write bogus reviews to earn credit and to promote (demote) the sales of product and stores.Although the occurrence of fake online reviews for products and services is deemed to have become an epidemic, little is known about the strategies used to write such bogus entries. Hence, this paper conducts an exploratory study to understand the process by which fake reviews are taken and propose a project which focuses on detecting fake and spam reviews by using different features with IP tracking and removes out the reviews which contains vulgar and curse words and make the e-commerce site fake review free online shopping center.

Keywords: fake reviews, spammers, sentiment analysis

INTRODUCTION

Consumers are usually inclined to seek information about products or services before making purchase decisions.Most of the online and professional review web sites provide service of writing reviews of the product, stores and services. Due to the importance of reviews to the product and stores, some fake reviewers emerge quickly to post fake reviews to promote sales or increase credit of the reviewer.These reviews are helpful for both individuals and business firms. These review systems motivate some people to enter their fake review to promote some products or downgrade some others. The main reason for this action is to make more profit by writing unfaithful reviews and false ratings .So in order to make products and services trustable, these fake opinions must be detected and removed. Detection techniques are used to discover fake reviews. Some fake reviews are written for malicious reasons to distort the reputation of businesses . Yet other fake reviews could be written with benign intentions to pass time, or gain status in the community . Conceivably, fake reviews could mislead consumers into making sub-standard purchase decisions. They tend to use the same review or slightly revised for different products. This duplication can be divided into four categories, (1) duplicates from the same customer id on the same product, (2) duplicates from different customer id on the same product, (3) duplicates from the same customer id on different products, (4) duplicate from different user id on different products.

LITERATURE SURVEY

The paper [1] proposes an algorithm for identifying the customer reviews. First, fake reviews are written to comprise short, catchy and succinct titles. In general, titles in reviews are known to play a significant role. Since they are displayed more conspicuously than review descriptions, they serve as relevance optimizers. Review titles are like headlines in newspapers, or taglines in advertisements.

The paper [2] Proposes On the practical front, this paper serves as an eye-opener for consumers suggesting that authenticity of reviews cannot be taken for granted. Consumers should becautious while reading reviews prior to making purchase decisions. Moreover, users should be honest and responsible while posting reviews. They should avoid submitting biased or misleading comments.

RELATED WORK

The proposed system will save their efforts and time by helping the users and business organizations identify spam from different opinions quickly and also help in purchasing their valuable products from a trustworthy site. To ensure credibility of the reviews posted on a platform, it is important to use a strong detecting model. Then a dictionary is made out of the current tokens and words in the given review which can be helpful in defining a sentiment index and weightage. In data mining the task of finding relationship among data is very important. Many

81

algorithm developed for this task, ie Apriori, fp-growth, eclat, relim etc. Data mining is a process of thoroughly scrutinizing huge store of data to discover patterns and trends that go beyond simple analysis. Data mining uses worldly-wise mathematical algorithms to fragment the data and evaluate possibility of future event. In fact if a potential customer gets a positive overall impression of a product by considering the present sentiments for that product, it is highly probable that he will actually purchase the product. Normally if the percentage of positive opinions is considerable, it is likely that the overall impression will be highly positive. Likewise, if the overall impression is negative, it is less imaginable that they don't buy the product.

EXPERIMENTAL SETUP

Module 1

The flowchart given below gives us an overview of how the system will work and how the each of the process will function. Flowcharts are used in designing and documenting complex processes or programs. Like other types of diagrams, they help visualize what is going on and thereby help the people to understand a process, and perhaps also find flaws, bottlenecks, and other less-obvious features within it.
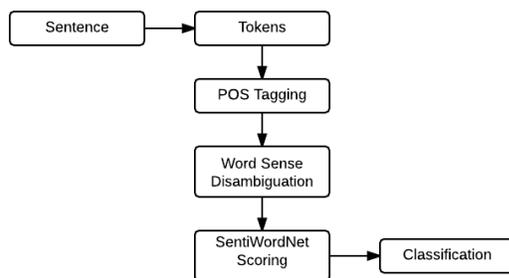
COMPONENTSOFTHEPROPOSEDSYSTEM

Data preprocessor

Data preprocessor component aims to collecting and cleaning the data before subsequent analysis. The data preprocessor module removes such non textual contents and gives the data in structured text format for using in opinion spam classification. The preprocessor removes noisy characters from the input documents.

Tokenization

Tokenization process splits the text into very simple units such as numbers, punctuation and words.

Sentence Splitter

Sentence splitter splits the whole text data into individual sentences. Usually Sentences begin with a capital letter which is the most identifiable sign for sentence splits. "." mark is considered as the sentence end, if it is not preceded by words such as Pvt., Ltd., etc. Some punctuation, such as ";", "!" And "?" are also used for sentence splitting

POS

It is also known as Parts of Speech (POS) tagging. Here every words are tagged into some index which helps in the future reference in the reviews. They are done so that we can easily handle words which are regularly used or are present in different reviews.

WSD

Word-sense disambiguation (**WSD**) is an open problem of natural language processing and ontology. WSD is identifying which sense of a word (i.e. meaning) is used in a sentence, when the word has multiple meanings. The solution to this problem impacts other computer-related writing, such as discourse, improving relevance of search engines, anaphora resolution, coherence, inference, *et cetera.*

SentiWordNet

SentiWordNet is the tool used for creating a dictionary of sentiment tokens. We divide the review sentence into individual tokens. The sentiment score shows a review's sentiment polarity. That is, the degree of how good or bad a review is. We calculated the sentiment score by incorporating the sentiment weight age of each word in the review.

Stop words Removal

Stop words such as 'a, an, is, the' are quite frequently seen in opinions but for classification there is no requirement of such words. So this component will remove such useless words for spam classification.

Sentiment analysis

Sentiment analysis is used to understand writers emotion. We define three list of words; positive vocabulary, negative vocabulary and neutral vocabulary, which consists of positive, negative and neutral words. Every review is passed to classifier which calculates the sentiment score of the reviews. Sentiment is a feeling, thought or judgement. It is also known as opinion mining, which studies people's feelings toward any entity. Internet

is full of resources with respect to sentiment information. From user's point of view, people can post their own matter on several social networking websites, such as forums, micro-blogs, etc. It is a flowchart that focuses our proposed model for categorization.

In Phase 2:

1) For negative phrase identification, an algorithm is proposed and implemented;

2) For sentiment score computation, a mathematical approach is proposed;

3) For sentiment polarity categorization, a feature called vector generation method is presented.

In Phase 3

1) Based on sentence- level and review- level, two sentiment polarity categorization experiments are performed respectively;

2) Performance of three classification models are evaluated and compared based on their experimental results.
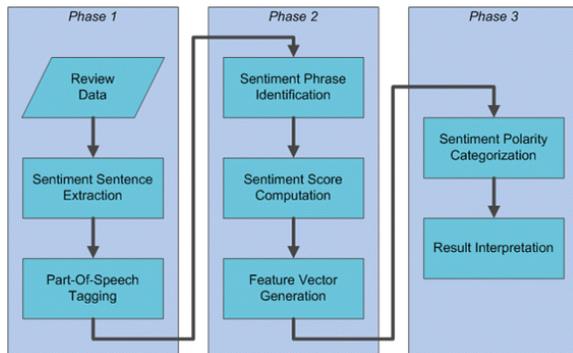
Module 2

This Flowchart says about the concept of IP restriction and access control. A symbol appearing in a particular "lane" is within the control of that organizational unit. This technique allows the author to locate the responsibility for performing an action or making a decision correctly, showing the responsibility of each organizational unit for different parts of a single process. The two most common types of boxes in a flowchart are:

☐ a processing step, usually called activity, and denoted as a rectangular box

☐ a decision, usually denoted as a diamond.
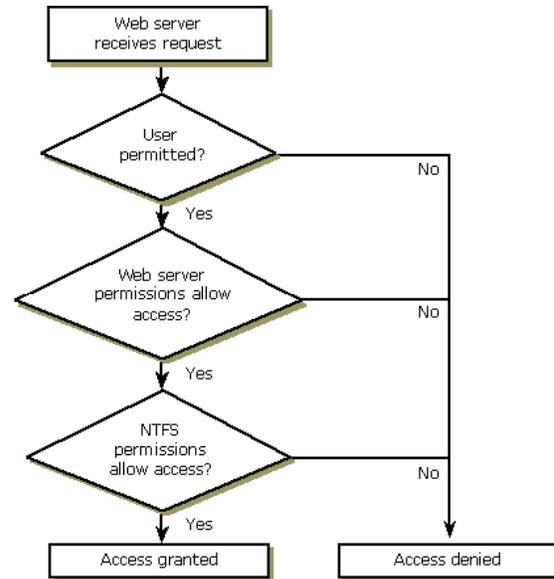
This project will analysis and then if any fake review is found from any IP address consistently then admin user can block that IP address. It also sends mail to user regarding blocked IP address. In this way it monitors the fake review made on any product. And user can be sure about the products availability on that application and reviews too with the help of Module 1.

CONCLUSION

There have been made several attempts for spam review detection till today. In this paper, we propose a general framework to detect spam reviews. Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, feelings, or emotions towards certain entities. In the literature survey there are many sophisticated methods explained which defines the sentiment analysis with respect to different aspects. Our application which will help the user to pay for the right product without any getting into any scams. Our application will do analysis and then post the genuine reviews on genuine product. And user can be sure about the products availability on that application and reviews too. The main objective of our work is to create a system which will detect spam and redundant reviews and to filter them so that user have a clear knowledge about the product. Aim of our project is to enhance customer satisfaction as well as to make online shopping reliable. The project will detect the fake reviews by deploying opinion mining algorithms and creating a word dictionary.

REFERENCES:

[1]Roth D, Zelenko D (1998) Part of speech tagging using a network of linear separators. In:Coling-Acl, The 17th International Conference on Computational Linguistics.

[2] Shashank Kumar Chauhan, Anupam Goel, Prafull Goel, Avishkar Chauhan and Mahendra K Gurve, "Research on product review analysis and spam review detection", 4th International Conference on Signal Processing and Integrated Networks(SPIN)

[3] Madhura N Hegde , Sanjeetha K Shetty , Sheikh Mohammed Anas , Varun K" FAKE PRODUCT REVIEW MONITORING" International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 06

[4] SP.Rajamohana, Dr.K.Umamaheshwari, M.Dharani, R.Vedackshya, "A survey on online review SPAM detection techniques", International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)

[5] Manleen Kaur Kohli, Shaheen Jamil Khan, Tanvi Mirashi, Suraj Gupta, "Fake Product Review Monitoring and Removal for Genuine Online Product Reviews Using Opinion Mining"

[6] Snehasish Banerjee Alton Y. K. Chua "Online Fake Product Review & Monitoring", Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore

[7] Sonu Liza Christopher and H A Rahulnath, "Review authenticity verification using supervised learning and reviewer personality traits", International Conference on Emerging Technological Trends (ICETT)

[8] RYK. Lau, SY. Liao, RCW. Kwok, K. Xu, Y. Xia, Y. Li, "Text mining and probabilistic language modeling for online review spamdetection", ACM Trans. on Management Information Systems(TMIS)

[9] A. Mukherjee, B. Liu, J. Wang, N. Glance, N. Jindal, "DetectingGroup Review Spam", International Conference on World Wide Web ACM

[10] Ruxi Yin, Hanshi Wang and Lizhen Liu, "Research of integrated algorithm establishment of spam detection system", 4th International Conference on Computer Science and Network Technology (ICCSNT)