IJCTA, 10(8), 2017, pp. 1-9 ISSN: 0974-5572

A Framework to Categorize Shill and Normal Reviews by Measuring it's Linguistic Features

Sarika S. Nalawade* and V.L. Kolhe**

ABSTRACT

Shill reviews detection has attracted significant attention from both business and research communities. Shill reviews are increasingly used to influence the reputation of products sold on websites in positive or negative manner. The spammers may create shill reviews which mislead readers to artificially promote or devalue some target products or services. Different methods which work according to linguistic features have been adopted and implemented effectively. Surprisingly, review manipulation was found on reputable e-commerce websites also. This is the reason why linguistic-feature based methods have gained more and more popularity. Lingual features of shill reviews are examined in this study and then a tool has been developed for extracting product features from the text used in the product review under analysis. Fake reviews, fake comments, fake blogs, fake social network postings and deceptive texts are some forms of shill reviews. By extracting linguistic features like informativeness, subjectivity and readability, an attempt is made to find difference between shill and normal reviews. On the basis of these three characteristics, hypotheses are formed and generalized. These hypotheses help to compare shill and normal reviews in analytical terms. Proposed work is for based on polarity of the text (positive or negative), as shill reviewer tend to use a definite polarity based on their intention, positive or negative.

Keywords: Informativeness; Linguistic characteristics; Readability; Reputation manipulation; Shill reviews; Subjectivity

1. INTRODUCTION

With the immense use of online reviews to purchase any service or product gives rise to fake reviews which misguides a consumer. Such review websites may permit a user to provide review for any type of service or product, for example, a restaurant, bar, hotel, Transportation Company (e.g. airline, train), shopping venue, spa and beauty service provider, financial service provider etc. Review websites are generally open for any registered or guest users to submit a review.

User generated content is mostly amorphous text, poorer quality, noisy, spam. The product information provided in reviews generally comes from actual users of the product. This data from actual product users helps rest of the users to reduce the risks related with buying products they have never used before [17]. However, the positive impact of item for consumption reviews on product sales provides a strong encouragement for sellers to manipulate reviews using deceptive reviews.

Fake or deceptive reviews are also called as shill reviews. Shill and shilling are the terms used about reputation manipulation. By enhancing this definition by specifying that a shill is a person who writes a review for a product without publishing the relationship between the seller and review writer.

* Department of Computer Engineering, D.Y. Patil College of Engineering, Pune Savitribai Phule Pune University, Pune, *Email: sarika.musale@gmail.com*

** Asst. Professor, Department of Computer Engineering, D.Y. Patil College of Engineering, Pune Savitribai Phule Pune University, Pune, *Email: vlkolhe@gmail.com*

Shill attacks are being more effective due to several reasons. To enumerate, important part of a product review is its overall rating. Numerical rating is one form of the review. Numerical rating of predefined aspects of the product or service is one option of expressing opinion. Therefore, thinly reviewed products, such as new products or specialized products, can benefit from shill attacks. Second, it is very simple to submit a review for a product. Usually, an account is necessary for a reviewer to submit a review, but the account registration process usually only requires the reviewer to have an email address, which can simply be obtained for free. Lastly, the identification of reviewers is often unnamed so reviewers do not have to be responsible for the content of their reviews. So, unlike reviews for sellers, many product review sites do not require reviewers to demonstrate product ownership prior to submitting a review.

Another option is short phrases summarizing pros and cons of product or service. General buyer seeks opinions from friends and family. Focus groups, opinion polls and surveys are some sources to get consumer feedback for which business spend a lot money.

As purchaser blindly trust on product reviews. To effectively detect shill reviews, there is a need to find the differences between shill reviews and normal reviews. A good understanding about the linguistic characteristics of fake reviews must be developed. In this paper, the linguistic characteristics of informativeness, subjectivity and readability in fake and normal reviews are analyzed using natural language processing (NLP) techniques. Data which contains fake reviews have been analyzed with this framework on the basis of informativeness, Subjectivity and Readability. To extend this work, negative reviews dataset is given as input to the system, because purchaser doesn't believe on negative reviews as compared to positive ones. In past, more importance was given to find out positive shill reviews, but in today's scenario it is equally important to find out negative shill reviews.

2. RELATED WORK

Many attempts have been made to prove the existence of review manipulation. Some other studies targeted duplicate reviews or doubtful reviews, not the direct cause of shill reviews. As a result, those studies categorized reviews from freely available websites as fake by detection of their level of duplication. As those studies did not use product features to separate shill and normal reviews, no methodology has been developed to identify and classify product features found in product reviews. This work will focus on differentiating shill and normal (authentic) reviews by using Description Based Feature Extraction Method (DFEM). DFEM uses 'informativeness', which is one of the linguistic features of text for calculating accuracy of reviews.

Following are some approaches for finding shill reviews.

2.1. Sentiment Analysis

As shill reviews are created to develop the positivity or negativity of a product or service, it may lead into positive or negative sentiment from online review. Sentiment classification aims to extract the text of written reviews of customers for certain products or services by classifying the reviews into positive or negative opinions according to the polarity of the review. Therefore, sentiment analysis can be employed as a tool for detecting shill reviews. This is done by computing the sentiment score of a review based on the sentiment scores of the terms used in the review. The sentiment of the review is defined to be equal to the sum of the sentiment scores for each term in the review.

Peng et al [13] used sentiment analysis to compute sentiment score from the natural language text by a shallow dependency parser. The relationship between sentiment score and spam reviews are discussed in further part.

Xiolong deng et al [16] has done further investigation on fake reviews on 'hype'. By human tagging of sentiment words, they have classified those words into four dimensions - service, overall attitude, taste and environment. The bayes classifier conducts sentiment analysis, and if the analysis result of four dimensions is same, then the review is defined as 'hype review'.

2.2. Linguistic features of product reviews

I. Informativeness: Quantity of product information provided in a review can be called as 'informativeness of the review'. Product information can be divided in following categories.

- *Official features:* An official feature is a noun or a noun phrase of the product which is included in the official product description. Official features are generally the product information that a consumer sees when reading the description of the product. An official feature is public information, normally provided by the company.
- *Unofficial features:* An unofficial feature is also a noun or a noun phrase[NP] about the product. This is not a part of product description provided by the manufacturer. It can be called as confidential information known only to the users of the product.

II. Readability: Readability is generally measured by the length of the text, the complexity of the words and number of sentences used in the review. Deceptive reviews contain more complex words, and making them less readable than genuine ones. Review length is the count of words in the given review. There are some index measures used to calculate readability of a given text in the review, such as Gunning-Fog Index, Coleman-Liau Index, Automated Readability Index and Flesch-Kincaid Grade Level. For a given text, the arithmetic mean of index measure of all sentences is called as 'Readability' of that text. Hence the general tool to classify a comprehensibility aimed features are number of words used in the text and 'mean readability' of the text.

III. Subjectivity: Subjectivity analysis is to classify a sentence into one of the two classes: objective and subjective. Positive and negative polarity of sentences can be extracted by measuring subjectivity of review. A subjective sentence gives very private information about the product and an objective sentence lists the features of the product. Spammers generally avoid statements of ownership because of lack of personal experiences. This specifies that shill reviewers will avoid subjective statements in their reviews because they have never actually owned the product. But, they are more likely to concentrate on describing the product. As normal reviewers have owned the product, they have experience using the product and will be free to express their feelings about the product. So, it can be assumed that normal reviews are expected to include more subjective sentences than shill reviews.

IV. Writing style: In order to reflect the opinions, spammers use their specific writing style to build sentences. To express an opinion about a specific product, reviewer uses a particular style. Writing style consists of the use of sentiment words, deceptive words, tenses as well as punctuations in reviews. Main feature of writing style used by a researcher is 'Stylometry'. This is scientific method used especially for security research as it helps to detect authorship of unknown documents.

Michael P. O'Mahony et al [11] addressed issues in the perspective of user generated product reviews. For easy purchase, product reviews have become an important asset to users that enables assessments of product quality. In particular, their focus was on features relating to the structure and readability of review texts, and examines the classification performance provided by these features.

Ee-Peng Lim et al [5] tried to find shill reviews generated by consumers. They made use of behaviors of review spammers by identifying their several characteristic. By using web based spammer evaluation software, they made a subset of highly suspicious reviewers for further processing.

Snehasish Banerjee et al [3] showed the difference between genuine and shill reviews in context of three textual features, like comprehensibility, informativeness and writing style. By collaborating multiple classification algorithms through polling, the analysis is done. Results verify that, reviews those are rich in nouns are expected to be genuine, whereas those rich in past tense, pronouns and articles are likely to be shill.

Jo Mackiewicz et al [20] stated that three characteristics of product reviews namely *Credibility, informativeness and readability*, positively affects review quality, and perceptions of quality is strongly influenced by the feature of informativeness, mainly a statement of the amount to which the product met the reviewer's expectations. These results imply that informativeness is the most important component of review quality perceived by users.

2.3. Machine Learning Techniques

Machine learning techniques were used frequently by past researchers to detect fake reviews [7]. Current research using supervised learning methods has been restricted to three learners: Logistic Regression (LR), Naïve Bayes (NB) and Support Vector Machine (SVM), even if there is a large number of machine learning algorithms (learners) available. Although SVM normally offered the best performance, it is rarely beaten by NB or LR, so it cannot be said as best learner.

Fan et al. [9] derivate a Statistical Opinion Analyzer (SOA) which extracts the polarity of online user reviews using NB classifier and frequency distribution. This framework makes it easy for a new consumer to buy a product and select manufacturer to increase the product's functionality. First, Reviews were crawled then pre-processed by GO tagger and inserted in SOA to find the positive and negative opinion probability with frequency distribution. This application gives promising results.

Tian et al. [10] devised a framework on Vietnamese reviews of mobile phones. By using HAC clustering and semi-SVMkNN classification synonym feature words were grouped. Opinion words along with weights have been used to extract feature words using pre-constructed adjective words and Viet Senti Word Net. Then, positive, negative and neutral polarities have been extracted, which is based on the weights and are used for opinion orientation.

Ott et al. [12] conducted a more current study of deceptive opinion spam by using the same data and framework as they used previous; on the other hand, they restricted their scope to n-gram based features and only used the SVM classifier since it outperformed Naïve Bayes in their previous work. An accuracy of approximately 86% is achieved by using unigram and bigram term frequency features when considering only reviews with negative sentiment.

3. IMPLEMENTATION DETAILS

To differentiate fake and normal review, it is required to measure the characteristics of shill reviews. There are methods to measure the features of a shill review. Figure 1 shows the block diagram of DFEM which is based in informativeness of a product feature.

3.1. Description-based feature extraction method

The count of official and unofficial features in a review defines its Informativeness. Following steps are followed to find Informativeness of a review.

Steps:

1. Collect the target product technical description.
2. Crawl to get all reviews of the target product, the technical description and reviews of all relevant products in the same category as product under study.
3. Pre-process the reviews of the target product for POS tagging.

4. Extract nouns and noun phrases from the reviews of the target product and compare them with those found in product technical description of the target product.
5. If a term used in review is also used in the product description, then it can be identified as an official feature.
6. The terms which do not appear in the product description, go through a filtering process that uses the technological description of other products in the same category to recognize which terms represent unofficial features of the product.

DFEM performance is calculated by following performance measures.

- $$\text{Recall} = \frac{TP}{TP+FN} \tag{1}$$
- $$\text{Precision} = \frac{TP}{TP+FP} \tag{2}$$
- $$F = 2 \cdot \frac{precision.recall}{precision+recall} \tag{3}$$

Readability is measured through following index measures:

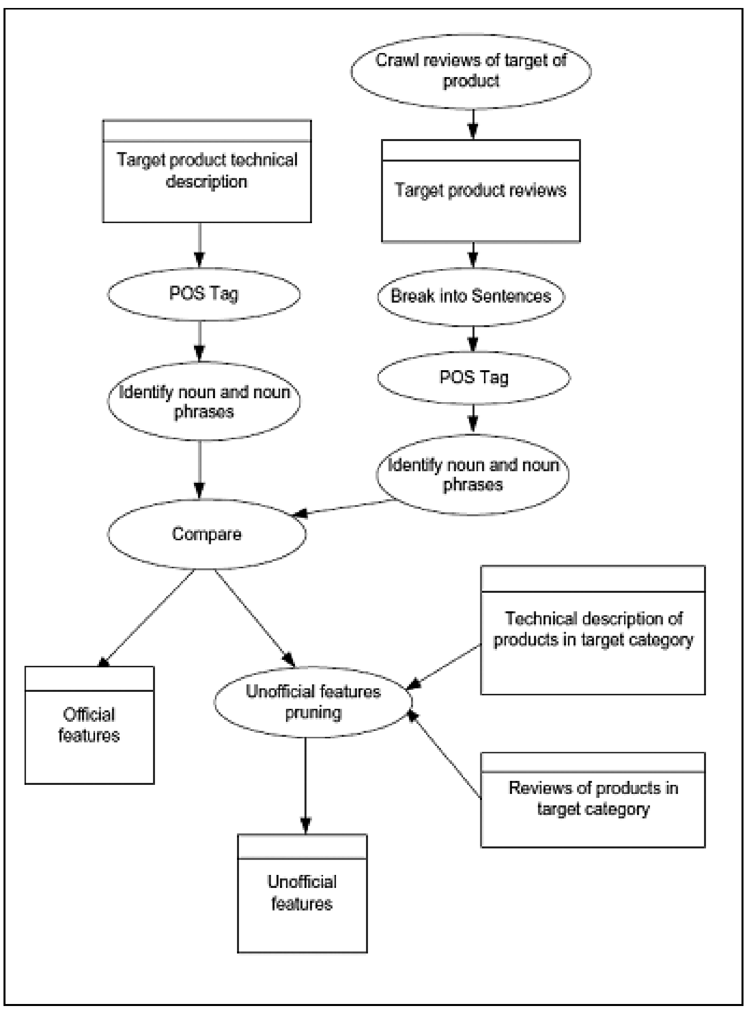

Figure 1: Description based feature extraction method.

1. The Fog Index: The value range of the Fog Index is from 1 to 12. A lower Fog Index means more readable text. The Fog Index of each review can be calculated as follow:

$$Fog = 0.4 * \left\{ \frac{words}{sentences} + 100 * \left(\frac{complex\_words}{N(words)} \right) \right\} \tag{4}$$

2. The Flesch Kincaid or Flesch Reading Ease Index: The value of this index is from 0 to 100, smaller scores indicating less readable text.

$$FK = 206.855 - 1.015 * \left(\frac{N(words)}{N(sentences)} \right) + 84.6 * \left(\frac{N(syllables)}{N(words)} \right) \tag{5}$$

3. The Automated Readability Index (ARI)

   The value of this index is from 1 to 12, number indicates the grade level education needed to understand the text. For example, ARI = 5 requires the reader to have fifth grade education to understand the text. ARI can be calculated as follow:

$$ARI = 4.71 * \left(\frac{N(characters)}{N(words)} \right) + 0.5 * \left(\frac{N(words)}{N(sentences)} \right) - 21.43 \tag{6}$$

4. The Colemon-Liau Index (CLI)

   The CLI ranges from 1 to 16 indicating the grade level education needed to understand the text.

$$\text{CLI} = 0.0588\text{L} - 0.296\text{S} - 15.8 \tag{7}$$

   L: number of characters per 100 words.

   S: number of sentences per 100 words

5. Simple measure of Gobbledygook (SMOG)

   A SMOG result also ranges from 1 to 12. SMOG is calculated as follow:

$$SMOG = 1.043 \sqrt{30 \frac{\text{Quantity of polysyllables}}{\text{Quantity of sentences}}} + 3.1291 \tag{8}$$

3.2. Dataset

The dataset taken for this study requested named as Cell Phone reviews. This dataset consists of reviews from Amazon. The data duration a period of 18 years, including 35 million reviews up to March 2013. Reviews consist of attributes like product and consumer information, ratings, and a plain text review.

3.3. Results

From cell phone dataset, first 100 reviews are extracted to evaluate the performance with recall, precision and harmonic mean measures. Another dataset is created which contains negative reviews only with same category and product description.

In figure 2, classification of features is done on positive reviews and negative reviews, results shows that negative reviews contains less no of official features compared to positive reviews. It also specifies that positive reviews contain fewer unofficial features. This concludes that negative reviews are more authentic than positive one for a given dataset.

Readablity is one of the effective measures to classify shill ans normal reviews. Fog index is performance measure to calculate readability value of reviews. Fig 3 shows the fog index values for various reviews. Readbale zone shows the reviews which can be read easily. By calculating fog index values for negative reviews, it shows that, 1 review is more readble than other 4 reviews.

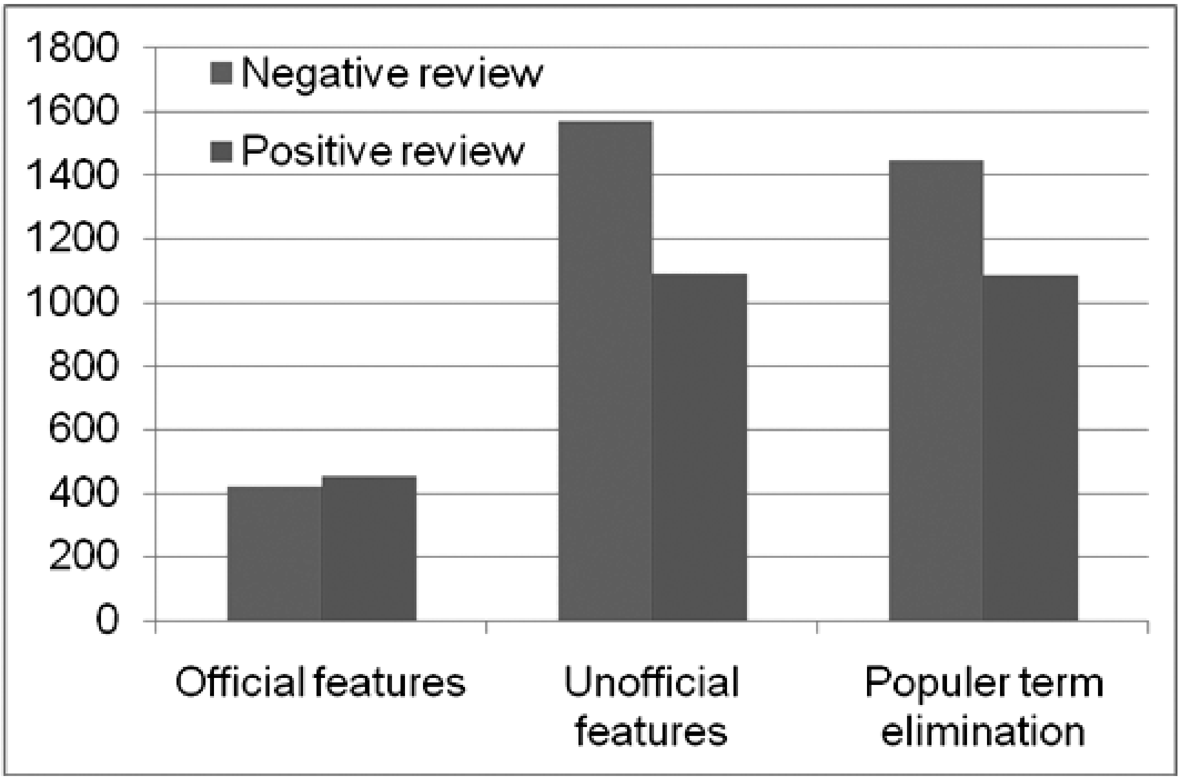

Figure 2: Performance of DFEM

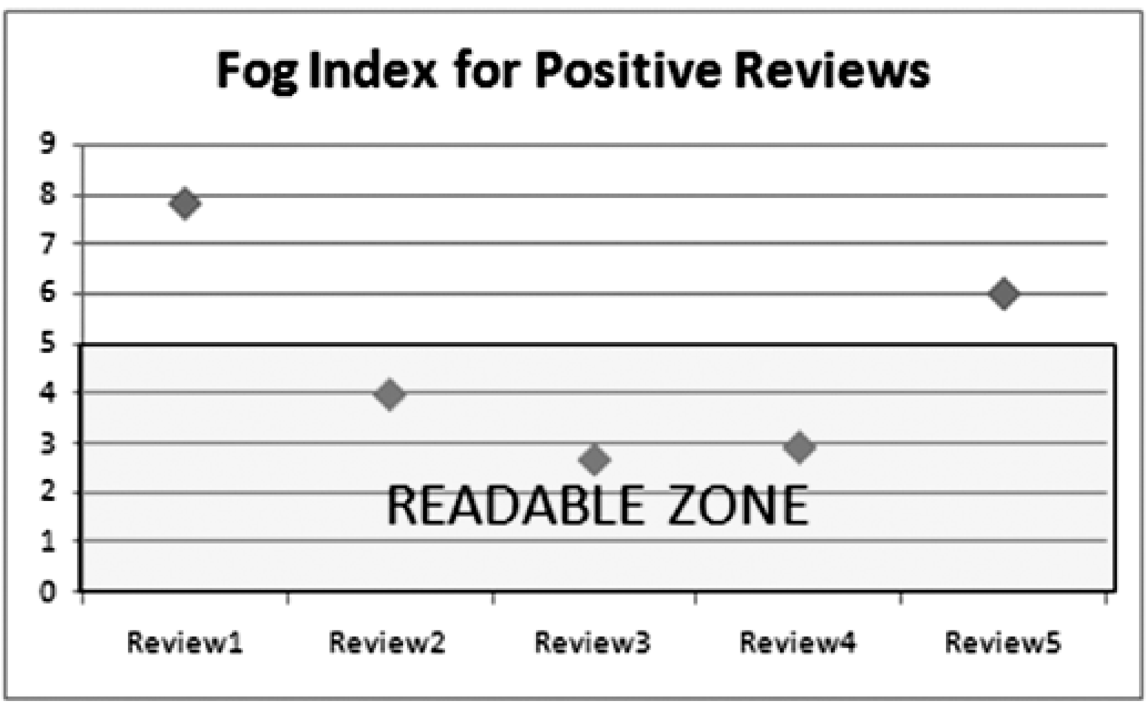

Figure 3: Readability performance for Positive reviews

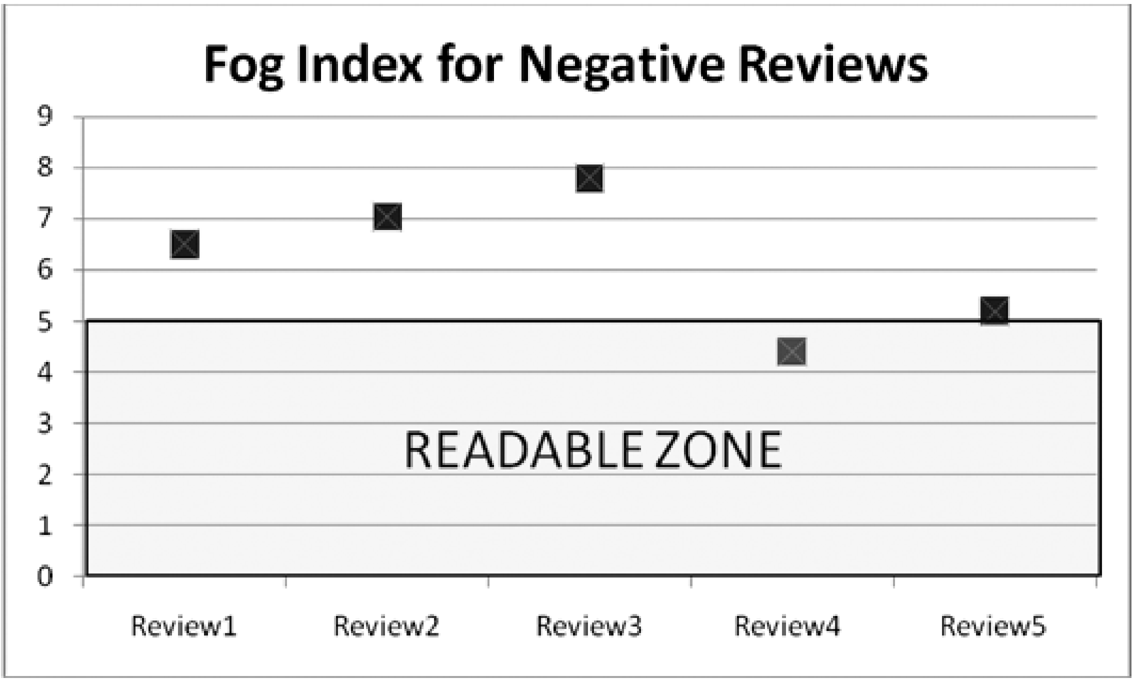

Figure 4: Readability performance for negative reviews

5. CONCLUSION AND FUTURE WORK

A general evidence of shill review is that, it is long enough and occurs frequently. Reason being, the spammer wants to grab attention of readers to official features of the target product or service. One can identify a shill review or reviewer based on the content of the review. Shill reviewer tend to use more objective features copied from the product/service specification sheet. On the other hand, a genuine or normal reviewer who also might have used the product/service by himself/herself tends to give more personal opinion hence being more subjective. As negative reviews are more likely to be shill, here for a given dataset, it showed that positive reviews are more fake than negative reviews.

Future work may focus on other methods which will measure other features like credibility, comprehensibility. Further there can some work done on identification of text polarity to identify shill reviews by using more representative dataset

ACKNOWLEDGMENT

We express our deepest gratitude to the college authorities for technical guidance and infrastructure. Lastly we wish to thank the researchers and reviewers for their contributions because of which we could complete this work.

REFERENCES

[1] Hiremath E. Patil A. Algur, S. and S. Shivashankar. “Spam Detection of Customer Reviews from Web Pages” In Proceedings of the 2nd International Conference on IT and Business Intelligence held in IMT Nagpur, 2011.

[2] E. Anderson and Simester. “Deceptive reviews: the influential tail. Working paper” In Sloan School of Management, MIT, Cambridge, MA, 2013

[3] David Bounie. “The Effect of Online Customer Reviews on Purchasing Decisions: the Case of Video Games” In Case study on purchasing decisions, 2005.

[4] Nitin Jindal Ee-Peng Lim, Viet-An Nguyen. “Detecting Product Review Spammers using Rating Behaviors” In CIKM ’10, October 26 to 30, 2010, Toronto, Ontario, Canada., pages Vol. 1, Page 309, October 26-30.

[5] Minqing Hu and Bing Liu. “Mining Opinion Features in Customer Reviews” In American Association for Artificial Intelligence, 2004.

[6] Li F, Yang Y. Huang, M. and X. Zhu. "Learning to Identify Review Spam" In Proceedings of the 22nd International Joint Conference on Artificial Intelligence ACL

[7] N. Jindal and B. Liu. "Review Spam Detection" In Proceedings of WWW ACM 2007, May 8–12, 2007, Banff, Alberta, Canada (Poster paper). Nov 2007.

[8] Barry O' Mahony, Michael P.; Smyth. "Using readability tests to predict helpful product reviews" In RIAO'10, 2010, Paris, France, 2010-04-28.

[9] Choi Y. Cardie C. Ott, M. and J.T. Hancock. "Finding Deceptive Opinion Spam by any Stretch of The Imagination" In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages Vol. 1, Page 309, 2011.

[10] Qingxi Peng and Ming Zhong. "Detecting Spam Review through Sentiment Analysis". In Journal Of Software, Vol. 9, No. 8, August 2014, page doi:10.4304/jsw.9.8.2065 to 2072, 2010-04-28.

[11] Ana Maria Popescu and Oren Etzioni. "Extracting Product Features and Opinions from Reviews" In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, page 339-346, October 2005.

[12] Ziqiong Zhang a Rob Law Qiang Ye a, b. "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches" In Expert Systems with Applications, page 6527-6535, 2009.

[13] Michael Brennan Sadia Afroz and Rachel Greenstadt. "Detecting Hoaxes, Frauds, and Deception in Writing Style Online" In Proceedings of the Conference on Web Search and Data Mining, 2012.

[14] Samaneh Moghaddam, Martin Ester, "Opinion Mining in Online Reviews: Recent Trends", Tutorial at 22nd International World Wide Web Conference, 13-17th May 2013.

[15] HaoWu Jingwei Zhang XiaolingWang Aoying Zhou Yuming Lin, Tao Zhu. "Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence" In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2014.

[16] Snehasish Banerjee, Alton Y. K. Chua "Let's Vote to Classify Authentic and Manipulative Online Reviews: The Role of Comprehensibility, Informativeness and Writing Style" Science and Information Conference 2015 July 28-30, 2015 | London, UK

[17] Jo Mackiewicz and Dave Yeats "Product Review Users' Perceptions of Review Quality: The Role of Credibility, Informativeness, and Readability" IEEE Transactions on Professional Communication, Vol. 57, No. 4, December 2014.

[18] Kolli Shivagangadhar, Sagar H, Sohan Sathyan, Vani, C.H Zhong "Fraud Detection in Online Reviews using Machine Learning Techniques" International Journal of Computational Engineering Research (IJCER) May – 2015

[19] Minqing Hu and Bing Liu. "Mining Opinion Features in Customer Reviews" In American Association for Artificial Intelligence, 2004.

[20] Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, "Survey of review spam detection using machine learning techniques" Crawford et al. Journal of Big Data (2015).