DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5473

ISSN No. 0976-5697

Volume 9, No. 1, January-February 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AN AUTOMATIC FRAMEWORK FOR DOCUMENT SPAM DETECTION USING ENHANCED CONTEXT FEATURE MATCHING

Yenuga Padma
Research Scholar in CSE,
Krishna University,
Machilipatnam (A.P).,India

Dr.Y.K.Sundara Krishna
Research Scholar in CSE Professor,Dept of CSE,
Krishna University,
Machilipatnam (A.P),India

Abstract**—** With the growth in the communication systems, opinions became the most used communication method in the corporates, research and education. Nevertheless, with the increasing popularity the challenge for all internet service providers is to keep matching the demand for bandwidth. The major challenge to keep the bandwidth up to the usage is dealing with the spam messages. A spam communication or review is something that the sender uses for promotion and for the received may be useless. Thus for the receiver the messages are mostly unimportant. The detection of the spam reviews cannot be done at the review server end and need to done at the receiver side. Failing in detecting the spam can easily overload the review communication channel and reduce the effective use of the bandwidth. A number of researchers are carried out in order to detect the spam messages by deploying the filters. The outcomes are partially satisfactory as most of the parallel researches have demonstrated the rejection of the documents based on the pre-defined keywords. Nonetheless, these methods are not satisfactory as the use of words for every review writer may vary. As a result influenced by certain keywords, the receiver may lose some important communications. Thus the demand of the modern research is to enhance the detection of the spam reviews by using enhanced techniques rather than only depending on the keywords. This work proposes a novel automated framework powered by machine learning technique to detect the keywords and improve the detection by deploying context detection methods. The major outcome of this work is to build and demonstrate an automated framework for review spam detection with review rejection filters. The work outcomes into a highly satisfactory detection rate and demonstrate a sustainable model.

Keywords: Spam Detection, Opinion Detection, Spam Filtration, Machine Learning, Feature Detection

I. INTRODUCTION

The notable review carried out by Yenuga Padma et al. [1] has demonstrated the types of the spam and barriers caused by those types. This review lays the demand for an automatic framework for opinion mining and detection of spam information specially the reviews. In modern times, people use web for everything, they use web to solve their questions, to find solutions of unsolved problems, to know about not so known products or services etc. They also use web, to know opinions of others before finalizing their decision on purchase of a new product or service. Positive reviews about products generally results in a purchase of a product and vice-versa. This reveals that opinions influence decision making of individuals and organizations.

However, the significant influence of opinions in decision making has also encouraged spammer and is also the reason behind the increasing number of opinion spams. Positive opinions can result in significant financial gains and/or fame for business, organizations and individuals. The negative opinions on some entities can damage their reputations.Deceptive opinions/fictitious reviews are

purposefully written to sound authentic and victimize readers.

The task of deceptive opinion spam detection can be modelled as binary classification problem with two classes, deceptive and truthful. Many of the previous studies on detecting deceptive opinions were based on methods that seek for duplicate reviews. The notable work by Nitin Jindal et al. [2] has demonstrated significant outcomes by deploying adaptive process for detection of spam reviews and information on the web. Some other researchers have also used meta-information such as the IP address of the reviewer or the average rating of the product, rather than the actual content of the review. The work by SihongXie et al. [3] has demonstrates a similar approach for the detection method. These works are highly criticised by parallel researchers as under the light of the IP masking and proxy servers, this methods are bound to fail.

Nevertheless, studies prior to this were not having access to any standard dataset and therefore their evaluations were based on some ad-hoc procedures as demonstrated by MyleOtt et al. [4] and Claire Cardie et al. [5]. Thus the demand of the modern research is to build an automatic framework for spam or fake review detection.

The rest of the work is furnished such as the outcomes from the parallel researches are realized in the Section – II, the proposed frame work is elaborated in the Section – III, the algorithm responsible for making the automated framework function is illustrated in the Section – IV, the comparative analysis based on the parallel researches on framework parameters and time complexity is disclosed in the Section – V, the results obtained from the framework is furnished in the Section – VI and the work presents the final conclusion from this work in Section – VII.

II. OUTCOME OF THE PARALLEL RESEARCH

Text content, behavioural analysis and supervised methods are used by many researchers to address the problem opinion spam detection. Jindal and Liu had first attempted the study of spam detection and had given two methods for spam detection based on duplicate detection and spam classification [6]. Jindal and Liu, in another study, identified opinion spam by detecting exact text duplicates in an Amazon.com dataset.

The parallel research methods listed out three types of duplicate positive reviews that were used as a spam [2]:

A. Duplicates from different user id on the same product

B. Duplicates from the same user id on different products and

C. Duplicates from different user id on different products.

MyleOtt et al. [4] proposed n-gram text categorization techniques to detect negative deceptive opinion spam with performance far surpassing that of human judges. Similar techniques for detecting positive deceptive opinion spam are proposed by Claire Cardie et al. [5].

Some studies that tried to trick better features to improve classifier performance used sentiment scores, product brand, and reviewer's profile attributes to train classifiers as demonstrated by Huang et al. [7]. Score computation based on behavioural heuristics, such as rating deviation as proposed by Nguyen et al. [8]. The study reported by Mukherjee et al. [9] focused on finding fraudulent reviewer groups by using frequent item set mining. Different stylistic, syntactical and lexical features describing opinions were identified [10]. They used support vector machine to learn a classifier based on these features of the opinions.

III.PROPOSED FRAMEWORK

The proposed framework is an automatic detector of fake or spam reviews on the popular web sites or the product pages. The component based framework is elaborated in this section [Figure -1].

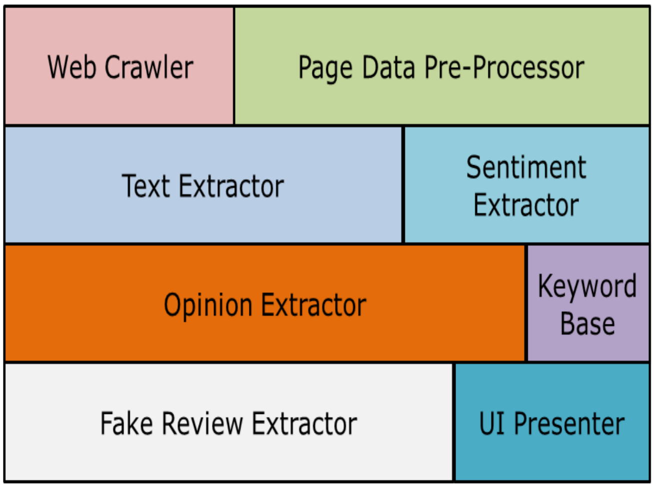

Fig. **Error! No sequence specified.** Proposed Novel Framework

A. Page Layout

The first component is the web crawler in this framework. The web crawler is responsible for crawling all the web links and detects the web links which readable. After listing the web links, the crawler algorithm finds the number of reviews per page. Then for all reviews the corpus is made read for analysis.

B. Page Data Pre-Processor

After the Web Crawler returns all corpus, the pre-processor component filters the unstructured data and fits into the structure which is predefined by this framework. The defined structure is listed here [Table – 1].

TABLE **ERROR! NO SEQUENCE SPECIFIED.** REVIEW DATA STRUCTURE FOR ANALYSIS

Parameter Name	Parameter Description
TimeStamp	The date and time for the review posted
Author	The screen name for the author
Review Text	The extracted text from the review
Positive Word List	Set of Positive Words Extracted
Negative Word List	Set of negative Words Extracted
Ratings	The numeric value for the product given.

C. Text Extractor

The Text Extractor fills in the data parameters for the Positive Word List and the Negative word list based on the associated English linguistics. The linguistics defines a set of positive and negative words along with the associated synonyms. The used collection is fabricated here with simplicity [Table – 2].

TABLE 2 EXTRACTOR COLLECTION – SAMPLE

Word	Type	Synonyms
AGILITY	Positive	NIMBLENESS, SUPPLENESS, ALERTNESS
BEST	Positive	FINEST, GREATEST, TOP
CAPABLY	Positive	PROFICIENTLY, SKILLFULLY, ABLY
DELIGHT	Positive	ENJOYMENT, PLEASURE, HAPPINESS
EXCITED	Positive	HAPPY, EAGER, MOTIVATED
GOODWILL	Positive	KINDNESS, FRIENDLINESS, FAVOR
ABRUPT	False – Positive (Spam)	SUDDEN, RAPID, HASTY
BASHFUL	Negative	RETIRING, MODEST, RETICENT
CAUSTIC	Negative	BURNING, SCATHING, CUTTING
DAUNT	False – Positive (Spam)	SCARE, OVERWHELM,

		FRIGHTEN
EXHAUSTS	False – Negative (Spam)	DRAINS, EXPENDS, DISSIPATES

D. *Sentiment Extractor*

The novelty of the framework is deployed majorly by this component. The extraction of the sentiment cannot be done only using the keywords. The sentiments are to be identified using the context of the sentence or the used key word. Based on the static information available in the linguist information set, the sentiment extractor highlights the reviews which are under the category of the false – positive or false – negative. These opinions are to be considered as spam as they give anomalous reviews. The sample extracted from the framework is fabricated here [Table – 3].

TABLE 3ANOMALOUS REVIEW DETECTION – SAMPLE

Word	Identified As	Spam
HASTY	False – Positive	Yes
OVERWHELM	False – Positive	Yes
EXPENDS	False – Negative	Yes

E. *Opinion Extractor*

The success of any model depends on the reduction of false identification and the false identifications can be reduced by using the validation models. The framework deploys a component as Opinion Extractor to match the opinion and the results of the Sentiment Extractor phase. If the extracted opinion is bad and the sentiment extractor result shows the review as positive, then the review should be marked as spam. This component is responsible for the validation.

F. *Keyword Base*

This is a static component in the framework consisting of widely accepted spam keywords. The keyword base identifies the results from the previous phase and justifies the results based on the type of keyword used.

G. *Fake Review Extractor*

The final resulting component in this framework is the Fake Review Extractor. This component based on the results from sentiment Extractor, Opinion Extractor and the Keyword Base defines the final results into the system.

H. *UI Presenter*

The UI Presenter is the presentation level component of this framework to demonstrate the final results to the end use.

IV.PROPOSED ALGORITHM

This section of the work elaborates about the algorithm that automates the framework.

```
Enable the Crawler by the product ID
Read the reviews online
     while (number of pages are not zero)     {
          read page for reviews;
          while(page contains review)  {
               read every review and store;
               Read review and map with EXTRACTOR COLLECTION framework;
               Extract Sentiment;
               if (context mismatch)          {
                    Mark Sentiment as False – Positive;
               }
               else
               {
                    Mark sentiment as False – Negative;
               }
               Extract rating as store as opinion;
               If (opinion is not same as sentiment)     {
                    mark the review as SPAM;
               }
               else
               {
                    mark the review as VALID;
               }
               mark page as read;
          }
     }
End Process;
```

The algorithm is analysed visually [Figure – 2].

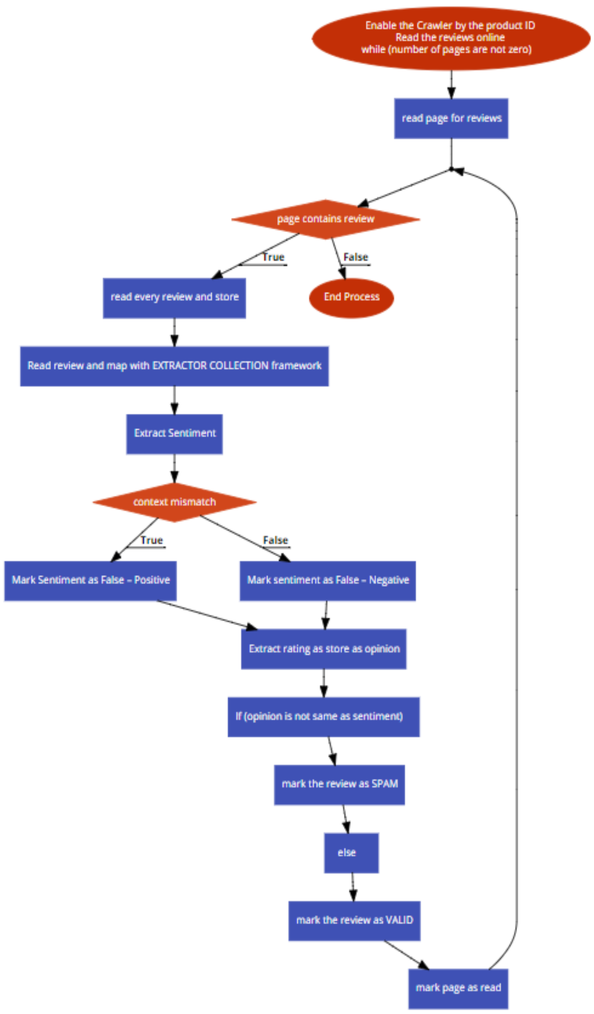

Fig. 2 Proposed Novel Automation Algorithm

V. COMPARATIVE ANALYSIS

In this section of the work the comparative analysis is carried out. The comparative analysis is based on two major factors as number of attributes used in the framework and the time complexity.

Framework Comparison

Firstly, the comparison on the frameworks are carried out and the findings are listed [Table – 4].

TABLE 4FRAMEWORK COMPARISON

Framework	Number of Parameters
Naïve Bayes	9378
SVM	82093
Proposed Algorithm	6

The number of parameters used in the proposed framework is significantly less. Nevertheless, the accuracy of the proposed framework is highly satisfactory.

Time Complexity Analysis

Secondly, the time complexity analysis is carried out and the findings are listed here [Table – 5]. The numbers of reviews per dataset are 92054.

TABLE 5TIME COMPLEXITY ANALYSIS

Framework	Time to build the model (ms)
Naïve Bayes	0.5
SVM	0.7
Proposed Algorithm	0.3

Thus the reduction in the time complexity is a clear indication of the improvements over existing methods.

VI. RESULTS AND DISCUSSION

The work is been tested with amazon customer review based on various product ids.

Firstly, the number of reviews extracted by this system is enlisted [Table – 6].

TABLE 6NUMBER OF PRODUCT REVIEWS EXTRACTED

Item Code	Item Name	Number of Reviews Extracted	Number of Actual Reviews
B075RWFCHB	Echo Plus with built-in Hub	5632	5632
B078Y4FLCL	Donkey Kong Country: Tropical Freeze - Nintendo Switch	0	0
B078PBR5C6	TAIR Wireless Bluetooth Headphone	1221	1221
B00923H7MA	Korg TM50BK Instrument Tuner and Metronome	882	882

Thus the pre-processing phase demonstrates significantly correct number of reviews fetching. The results are also analysed visually [Figure – 3].

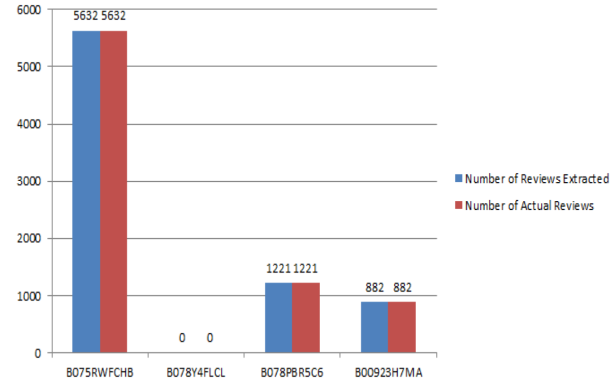

Fig. 3 Review Extraction

Further the detection of the negative, positive or the false positive or the false negative reviews are identified [Table – 7].

TABLE 7REVIEW ANALYSIS

Item Code	Positive	Negative	False Positive	False Negative
B078Y4FLCL	0	0	0	0
B078PBR5C6	1196	12	10	3
B00923H7MA	643	26	209	4

The results are visualized graphically [Figure – 4].

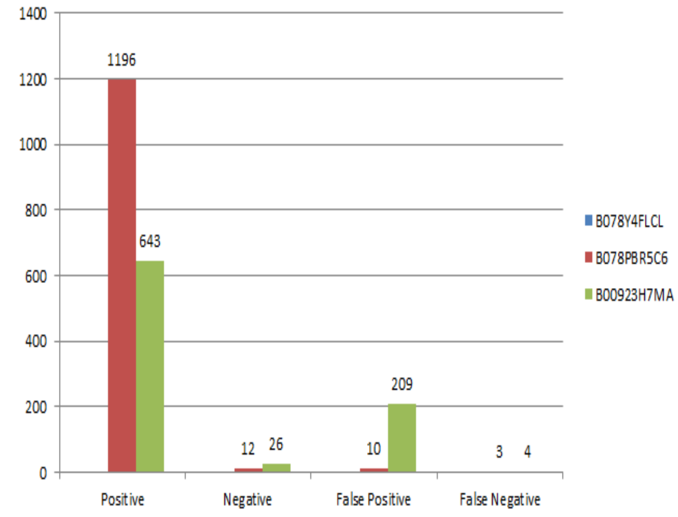

Fig.4 Sentiment Extraction

Finally, the work furnishes the spam detection results [Table – 8].

TABLE 8 REVIEW ANALYSIS

Model Name	False Positive	False Negative	Identified as Spam
Naïve Bayes	219	7	101
SVM	219	7	198
Proposed Algorithm	219	7	226

The results are visualized graphically [Figure – 5].

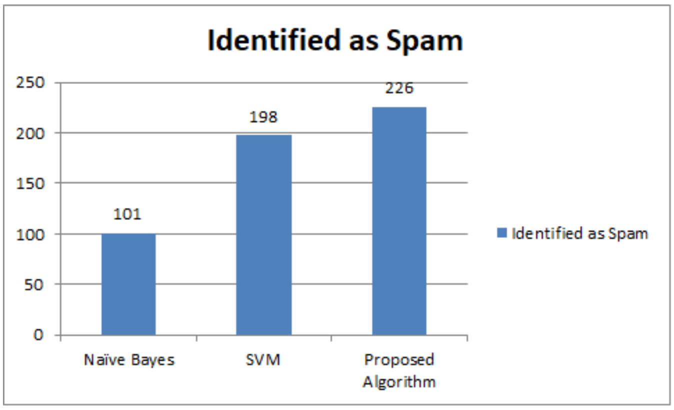

Fig. 5 Spam Review Detection

Also the accuracy of the models are analysed [Table – 9].

TABLE 9ACCURACY OF SPAM DETECTION

Model Name	Total Number of Spam Reviews	Identified as Spam	Accuracy (%)
Naïve Bayes	226	101	44.69
SVM	226	198	87.61
Proposed Algorithm	226	226	100.00

The results are analysed visually [Figure – 6].

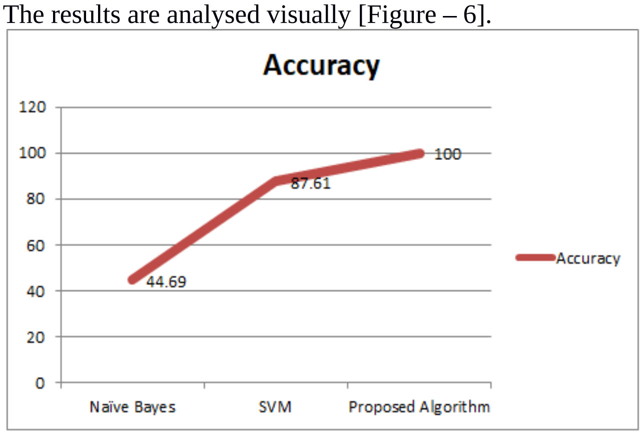

Fig. 6 Spam Detection Accuracy

Thus it is natural to understand that, the accuracy of the proposed algorithm and the framework is extremely satisfactory. Thus in the light of the reviews, results and comparative analysis, this work presents the conclusion in the next section.

VII. CONCLUSION

The growth in the online review systems influences many factors for the consumers. Any negative review or any positive reviews can complete change the perspective of the reader. Thus the review systems are to be considered with high importance and need to be validated. The spam reviews which are available online can destroy the branding completely. Thus this work deploys an automatic framework to validate the reviews and mark the spam reviews. The proposed framework enables the consumer or the reader to detect and eliminate the spam reviews completely. The work demonstrates a highly reliable framework with 100% extraction rate and 100% accuracy of the detection of fake or spam reviews. The highly satisfactory result is obtained due to the extraction of words justified by sentiment and validated by opinions. The outcome of the work is justified to make the online review system better for the world by reducing the negative influences by spammers.

VIII. REFERENCES

[1] Yenuga Padma, Dr. Y.K Sundara Krishna, “A literature review on Opinion spam detection and its approaches”, Vol. 15, Issue 6, June 2017, pp. 351-356, IJCSIS.

[2] Nitin Jindal, Bing Liu,” Opinion Spam and Analysis”, ACM Proceedings of the International Conference on Web search and web data mining, 2008, pp. 219-229.

[3] SihongXie, Guan Wang, Shuyang Lin, Philip S. Yu “Review spam detection via time series pattern discovery”, ACM Proceedings of the 21st international conference companion on WWW '12, 2012, pp. 635-636.

[4] MyleOtt, Yejin Choi, Claire Cardie, Jeffrey T. Hancock, “Finding deceptive opinion spam by any stretch of imagination”, Volume 1, 2011, ACM Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 309-319.

[5] Ott, Myle, Claire Cardie, and Jeffrey T. Hancock. "Negative deceptive opinion spam", Proc. NAACL-HLT, 2013, pp- 497–501.

[6] Nitin Jindal, Bing Liu ,”Review Spam Detection”, ACM Proc., 16th International conference on World Wide Web, 2007,pp. 1189-1190.

[7] Fangtao Li, F.; Huang, M.; Yang, Y.; and Zhu, X, “ Learning to Identify Review Spam” Proc. of Twenty-Second International Joint Conference on Artificial Intelligence, 2011,pp. 2488-2493.

[8] Lim, E. P., Nguyen, V. A., Jindal, N., Liu, B., & Lauw, H. W., “Detecting product review spammers using rating behaviours”, *Proc., International Conference on Information and Knowledge Management, 2010, pp*. 939-948.

[9] Mukherjee, A.; Liu, B.; and Glance, N. S.,“Spotting fake reviewer groups in consumer reviews”, Proc. 21st international conference on WWW '12 , 2012,pp. 191-200.

[10] Raymond Y. K. Lau, Stephen Shaoyi Liao, Ron Chi-Wai Kwok, Kaiquan Xu, Yunqing Xia, Yuefeng Li,” Text mining and probabilistic language modeling for online review spam detection” ACM Transactions on Management Information Systems (TMIS), Volume 2 Issue 4, December 2011, doi: 10.1145/2070710.2070716.

