ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal
Regular Issue, Vol. 8 N. 2 (2019), 61-71
eISSN: 2255-2863
DOI: http://dx.doi.org/10.14201/ADCAIJ2019826171

Detecting Spam Review through Spammer's Behavior Analysis

Naveed Hussain[a], Hamid Turab Mirza[a] and Ibrar Hussain[b]

[a] Department of Computer Science, COMSATS University Islamabad (Lahore Campus), Pakistan
[b] Department of Software Engineering, The University of Lahore, Lahore, Pakistan
fa15-pcs-002@cuilahore.edu.pk , drturab@cuilahore.edu.pk, ibrar.hussain@cs.uol.edu.pk

KEYWORD

Online Product Reviews; Spam Reviews; Spam Review Detection; Opinion Spam; Customer Reviews; Spammer Behavioral features

ABSTRACT

Online reviews about the purchase of a product or services provided have become the main source of user opinions. To gain profit or fame usually spam reviews are written to promote or demote some target products or services. This practice is known as review spamming. In the last few years, different methods have been suggested to solve the problem of review spamming but there is still a need to introduce new spam review detection method to improve accuracy results. In this work, researchers have studied six different spammer behavioral features and analyzed the proposed spam review detection method using weight method. An experimental evaluation was conducted on a benchmark dataset and achieved 84.5% accuracy.

1. Introduction

Customer reviews have become the main source to collect different opinions about products and services. These affect the daily life decisions and professional activities: e.g., which restaurant is good, which car to purchase, which product to buy and which doctor to consult. Spam reviews may directly interpret financial advantages and losses for a company (Hussain *et al.*, 2019; Bajaj *et al.*, 2017). Large number of favourable reviews about product and services may attract more customers whereas negative reviews are often a reason of decline in the sale. This, unfortunately, gives a strong incentive for opinion spamming which refers to illegal human activities (e.g., writing spam reviews and giving false ratings). Opinion spamming gives a wrong impression to customers by promoting/demoting certain entities (e.g., products and businesses) (Biradar *et al.*, 2017).

Most customers purchase products, which contain spam opinion and they could be disappointed after purchasing such products. Therefore, it is very important to identify and highlight the spam reviews. Now a days, spam review detection is a very challenging task and many researchers are working on it (Mukherjee *et al.*, 2013). Opinion sharing sites are starting to gain competitive advantages for business. However, the main drawback of opinion through opinion sharing sites is that anyone from anywhere in the world can post reviews about the product/service without any limits. A businessman can hire spammers to promote the product or support their sale (Fusilier *et al.*, 2015).

Since review spam can financially affect businesses and cause a sense of mistrust in public, therefore, this problem has recently attracted the consideration of media and governments. In the same context, media news from the New York Times and BBC have reported that "Nowadays spam reviews are becoming very common on the web sites and recently a photography company was subjected to hundreds of consumer fake reviews" (Ong *et al.*, 2014). Hence, detecting spam reviews and opinions are emerging as an important research area.

Commercial review hosting sites e.g. Yelp[1] and Amazon[2] have already put through some progress in detecting spam reviews (Dematis *et al.*, 2018). In 2008, the first research about opinion spamming was introduced by Jindal et.al. (Jindal *et al.*, 2008). Following this research, developments to find spam reviews have made significant progress and several different dimensions have been explored, such as detecting individual spammer (Liu *et al.*, 2012) and group spammers (Zhou*et al.*, 2018).

Review spam can be related to email and web spam, but it requires different techniques for its identification. The Web spam is used to attract people by manipulating the content of the page so that the web page will be ranked high by the search engine. Email spam is mainly used for ads. Spam reviews, on the other hand, are very different as they give the wrong opinion about the product/service and it is very difficult to detect spam reviews manually. Therefore, existing web spam or email spam techniques (Chakraborty *et al.*, 2016) are not suitable for spam review detection.

Spam review detection is a challenging task as no one knows exactly the amount of spam in existence. Spam review looks like a normal review until one applies different spammer behavioral features and linguistic features to identify the spammer and spam reviews respectively. Therefore, there is still a need to employ different spam review detection techniques to accurately filter out spam or not spam reviews. This research makes the following main contributions:

1. Introduced new spammer behavioral features such as the maximum number of reviews, activity window, review count and the ratio of the first review that has improved the accuracy of the proposed spam review detection method.
2. Analysed the proposed spam review detection method using the weight method to identify spam reviews.

The remainder of the paper is organized as follows: In Section 2, we briefly cover the state-of-the-art techniques related to spam review identification domain. Section 3 describes about the review dataset. Section 4 describes the proposed methodology for spotting spam reviews. In Section 5, an experimental study is conducted to demonstrate the effectiveness of the proposed method. Finally, Section 6 concludes the work.

2. Related Works

To tackle the problem of spam reviews, a variety of different spam review detection methods have been explored. This study reviews the literature from spam review detection using spammer behavioural features analysis method (unsupervised learning). This research also aims to determine the contribution by relating this latest work with prior studies.

A study (Mukherjee *et al.*, 2013) developed a spam review detection method using the clustering technique by modelling spamicity of the author that separates spammer and not spammer clusters. Similarly, Heydari et al. (Heydari *et al.*, 2016) have proposed a model dependent upon features such as time series on Amazon real-world dataset.

In another existing study (Kc *et al.*, 2016), authors offered a text mining model by using an unsupervised approach and features rely upon time integration among multiple time durations. In addition, this model integrates with the semantic language model for spotting spam reviews and used the Yelp dataset[3].

Another related study (Li *et al.*, 2017) have suggested that the author's spamicity unsupervised model is based on features such as review posting rate and temporal pattern. This proposed model produced two clusters: spammers and truthful user and dataset were crawled from Chinese website Dianping[4] to train the proposed

1 www.yelp.com
2 www.amazon.com
3 www.yelp.com/dataset
4 www.dianping.com

model. In another study (Dematis *et al.*, 2018), authors have observed a network model for spam review detection. In their work, the correlation among users and products is captured and the algorithm is used to recognize the spam reviews.

The study (Kaghazgaran *et al.*, 2017) identified some spammer behavioral features; like the length of the reviews, rating distribution, and the reviews burstiness and highlighted that these features can reveal the spammers and spam review. It may also yield clues as to which reviewers are fraudulent. In another study (Viviani *et al.*, 2017), the authors focused on the aggregation process with the aim of finding a veracity score with respect to each single veracity feature. They evaluate their method of Yelp dataset by using aggregation schemes.

Another related study (Wang *et al.*, 2018) introduced an LDA-based group spamming detection approach based on the mature LDA model and SCAN algorithm. They adapted the LDA algorithm for review spammer detection by boosting the count of suspicious reviewers according to review time burstiness and the rating score. The method only uses review time and rating score data. It did not consider review text.

In another study (Kaghazgaran *et al.*, 2018), the authors proposed a neighbourhood-based approach to detect spam groups. They monitor tasks on sites like Rapid Workers5 to uncover fraudulent reviewers on sites like Amazon. They proposed Twofaced framework for identifying online spam groups. The main purpose of their research was to: (i) find the locality of suspiciousness within the graph through a random walk to find suspicious users who tend to make cluster, and (ii) exploit the structure of the graph around suspicious users to uncover campaign network structures for identifying fraudulent reviewers who are distant in the graph.

Based on the gaps identified in the reviewed literature, the current study proposed a spam review detection model using spammer behaviour analysis. It introduces a number of spammer behavioral features such as a maximum number of reviews, activity window, review count, the ratio of the first review to identify spammers. The proposed model is based on weight method which identify the spam reviews.

3. Dataset

The study utilized car reviews dataset[6] (Ganesan *et al.*, 2012) collected from Edmunds[7]. The dataset contains 42,288 reviews about cars for the model-year 2007, 2008, and 2009. There are about 140-250 cars for each model year. The extracted fields include author names, dates, favourites, the full textual reviews and rating used as relevance judgments. The complete summary about dataset is presented in Table 1.

The real world collected dataset was noisy and contained null values attribute, so the dataset was pre-processed to improve the accuracy of the proposed system and after prepossessing 35,290 reviews were used to analyze the proposed method.

Table 1: Statistics about car reviews dataset

Year	Reviews
2007	18,903
2008	15,438
2009	7,947
Total	42,288

5 www.mturk.com
6 kavita-ganesan.com/entity-ranking-data
7 www.edmunds.com

4. Spammer Behavioral Features

This section explains the proposed spammer behavioral features and the normalized value (0-1) of each spammer behavioral feature is calculated. Moreover, these normalized values are used as input for proposed spam review detection method. Since a spammer may be identified by behavioral features. So, there is no need for the labelled dataset (Hazim *et al.*, 2018). Notations used in this section are listed in Table 2.

Table 2: List of notions

Variables	Description
a	Author 'a'
r	Review 'r'
r_a	Review 'r' by the author 'a'
R_a	Set of all reviews of an author 'a'
t	Time of current review 'r'
L_a, F_a	Time of the last review of the author 'a', Time to the first review of the author 'a'
*(r)	Rating of review 'r'
$MaxRev(a)$	Maximum number of reviews by the author 'a'

i. Author Content Similarity (CS)-F1:

It is a time-consuming activity for the spammer to write a new spam review. Spammers often post those reviews which are identical or near to identical of their previous reviews. Thus, researchers can calculate the author's review content similarity to identified similar reviews by using the subsequent equation.

$$F_1(a) = \max_{x}^{y} \; [cosine(r_i, r_x)] \; where \;\; r_i, r_x \in R_a, x < y \tag{1}$$

In Eq. (1), r_i the notion is used to represent the current review. The total number of reviews written by that author is represented by y. In this study, researchers used a cosine similarity function to calculate the similarity of each author review with the previous reviews.

ii. Maximum Number of Reviews (MNR)-F2:

It is a common activity of the spammer that posts too many reviews in a single day. Therefore, it is assumed as the irregular behaviour of the reviewer. Eq. (2), is used to calculate the ratio of the total number of reviews by an author *a* by the maximum number of reviews posted by that author in previous days.

$$F_2(a) = \frac{MaxRev\,(a)}{max(MaxRev(a))} \tag{2}$$

iii. Activity Window (AW)-F3:

It has been observed by the literature review that spammers are not regular user/member of the system (Zhou *et al.*, 2017). Moreover, genuine reviewers regularly post reviews about products or services. Therefore, it is an important spammer behavioral features to identify the irregular users of the system. Eq. (3) is used to calculate the difference between timestamps of first and last review of an author to find out a number of active days a reviewer remains on the system. (Threshold, X = (0,45]).

$$F_3(a) = \begin{cases} 1, L_a - F_a < X \\ 0, \ \ otherwise \end{cases} \tag{3}$$

iv. Review Count (RC)-F4:

As discussed above, spammers are not long-term members of the system. Therefore, the spammer is more likely to have a lesser number of reviews than that of the genuine reviewer. The dataset was examined to find out the number of reviews a spammer usually writes. Therefore, threshold value X=5 is used to identify spammers or genuine reviewers. The Eq. (4) is used to calculate the value of the review count.

$$F_4(a) = \begin{cases} 1, \ \ |R_a| \ < \ X \\ 0, \ \ otherwise \end{cases} \tag{4}$$

v. Ratio of First Review (FR)-F5:

Early reviews about the product/service increase the impact of sale and people also rely on early reviews. Therefore, spammer tries to post reviews as early as possible to control the opinion about product/service. Eq. (5) is used to calculate the ratio of the first review of an author by a total number of reviews of that author.

$$F_5(a) = \frac{\sum_{x=1}^{|R_a|} |\{r_x \ \in \ R_a\} \cap (r_x isafirstreview)|}{|R_a|} \tag{5}$$

vi. Rating Deviation (RD)-F6:

The spammer is usually giving wrong projection about the product/service either in a positive sense or negative sense. It was observed by the literature review that spammer rating mostly deviates from the average rating of the reviewer. The Eq. (6) and Eq. (7) are used to calculate the normalized value of rating deviation.

$$MEAN_r = \frac{\sum_{x=1}^{|R_a|} |\star \ (r_x)|}{|R_a|} \tag{6}$$

$$F_6(r) = \frac{|\star \ (r_a) - MEAN_r|}{4} \tag{7}$$

4.1. Spam Review Detection using Weight Method

Based on six different spammer behavioral features suggested in Section 4, the weight method is used to find the spam reviews. Considering the review dataset, each review is assigned a label from the set L = {L_{normal}, L_{spam}}. L_{normal} is used for normal reviews and L_{spam} is used for spam reviews. The proposed threshold-based spam review detection method uses the calculated normalized values of spammer behavioral features from F1-F6 (Section 4). The following Eq. (8) is used to calculate the spam score of each review:

$$Spam\ Score \ = \frac{(\alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3 + \alpha_4 F_4 + \alpha_5 F_5 + \alpha_6 F_6)}{\sum_{K=1}^{6} \alpha_k} \tag{8}$$

Experiments are performed to investigate the contribution of each spammer behavioral features and how to adjust the weight (Asadi *et al.*, 2015; Pudaruth *et al.*, 2018) of each behavioral feature is also described in section 5. Moreover, spam review is highlighted with a predetermined threshold, such as $\tau = 0.5$, 0.55 and 0.6 and Eq. (9) is used to label each review from {L_{normal}, L_{spam}} where *i* is the review number. The proposed framework is presented in Figure 1.

$$L_{r[i]} = \begin{cases} L_{normal} & Score_{\ r[i]} < \tau \\ L_{spam} & Score_{\ r[i]} > \tau \end{cases} \tag{9}$$

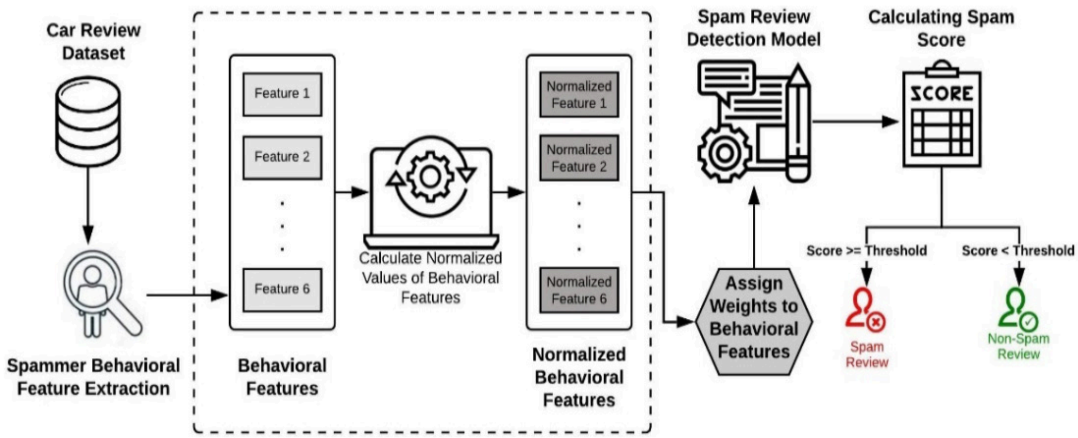


Figure 1: A proposed framework for spam review detection using weight method.

Algorithm 1: Spam review detection using weight method

```
Input: review rj
Output: Spam or Not-Spam
   for each review rj in review dataset do
      // behavior features (F1, F2, F3, F4, F5, F6)
      for each behavior feature Fi calculate normalize value do
         calculate normalize value of Fi
      end for
      for each behavior feature Fi do
         assign weight
      end for
      if Spam Score > Threshold then
         label rj ←— Spam
      else
         label rj ←— Not-Spam
      end if
   end for
```

Figure 2: Algorithm of proposed spam review detection model using weight method.

5. Experimental Evaluation

In this study, spam review detection method using spammer behavioral features has two phases: (1) Calculate the normalized value (0-1) of each spammer behavioral feature. Based on these normalized values, calculate the spam score using the weight method. (2) Evaluate the performance of a proposed spam review detection using weight method by using different threshold values such as 0.5, 0.55 and 0.6. The performance of the proposed spam review detection method is evaluated by precision, recall, f-measure and accuracy. Moreover, Support Vector Machine classifier used for training and testing the proposed model. Spam review detection using the weight method uses two different sets of parameters: The feature weight and the threshold value. Table 3 shows the efficiency of the threshold-based method with the combinations of different threshold values and different weight combination of behavioral features. For example, the "211222" means the weight of F1, F4, F5 and F6 are 2 and other features F2 and F3 are 1. Table 3 shows that we get the maximum accuracy value when the threshold value is τ =0.60 and the feature weight setting is "122221".

The value of threshold τ can be adjusted to achieve the desired efficiency for different applications, if we want to filter maximum spam reviews as possible e.g. for a critical product such as medicine or other health-related services then we should use the value of threshold to be relatively greater. If we are not stricter about spam reviews for any product/services such as movies, leisure or other such products then we may set small threshold values. Table 3 clearly shows that if the threshold value τ is maximum than accuracy will be higher.

Table 3: The effectiveness of the threshold-based detection method

Threshold	Feature Option	precision_s	recall_s	F − measure	Accuracy
$\tau = 0.50$	111111	0.752	0.731	0.745	0.732
	211112	0.749	0.749	0.752	0.741
	121111	0.759	0.739	0.759	0.750
	122111	0.769	0.756	0.765	0.762
	122211	0.764	0.760	0.769	0.768
	122221	0.779	0.775	0.770	0.771
$\tau = 0.55$	111111	0.753	0.749	0.753	0.748
	211112	0.761	0.751	0.763	0.758
	121111	0.762	0.760	0.762	0.759
	122111	0.775	0.765	0.769	0.762
	122211	0.781	0.776	0.779	0.775
	122221	0.791	0.781	0.792	0.788
$\tau = 0.60$	111111	0.778	0.752	0.769	0.761
	211112	0.794	0.771	0.784	0.774
	121111	0.801	0.799	0.798	0.794
	122111	0.825	0.814	0.821	0.812
	122211	0.837	0.824	0.834	0.821
	122221	0.840	0.831	0.834	**0.845**

Figure 3 shows that out of total 35,290 car reviews, τ =0.50 detected only 4,211 spam reviews, τ =0.55 detected 6,356 spam reviews and after applying τ =0.60 detected 9,231 spam reviews. This shows that a larger value of the threshold can detect a maximum number of spam reviews.

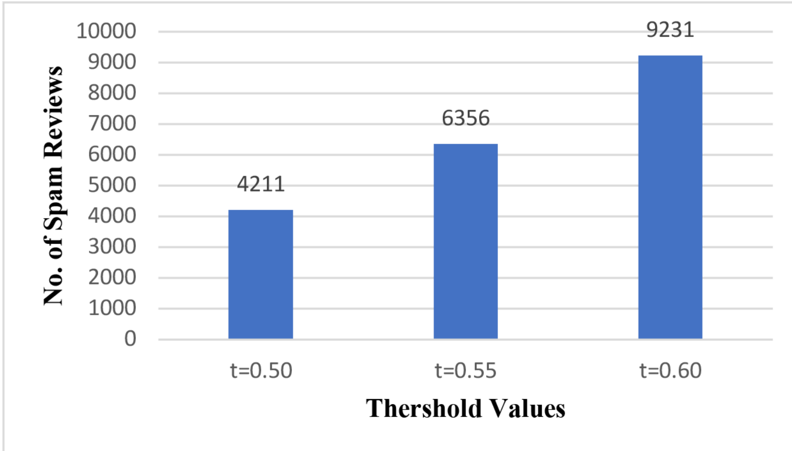


Figure 3: Comparison of different threshold values.

5.1. Reviewers Behaviour Analysis

This section studies the behaviors of reviewers in detail. For the purpose of this study, spammers and non-spammers have already been identified in the dataset using the above-explained model. The reviewer's profile is studied under the following dimensions using the Cumulative Distribution Function (CDF).

i. Content Similarity (CS)-F1:

Figure 4(a) shows that almost 50% of spammers have a similarity index between 0.1 to 1. However, the rest 50% have exactly 1 similarity index of their reviews which means that almost half of the spammers copy their previously written reviews instead of writing a new review. It can also be observed that non-spammers also write similar reviews, almost 60-70% of non-spammers have little similarity index of 0.5 approximately.

ii. Maximum Number of Reviews (MNR)-F2:

Writing too many reviews in a single day is not considered as normal behaviour. CDF in Figure 4(b) shows that only 5-8% of spammers have a lesser ratio of posting reviews in a single day. However, the rest of the spammers show that they tend to write more and more reviews in a day. For non-spammers, they have a moderate reviewing rate. Approximately 60% never wrote more than a single review per day and the ratio of their posted reviews increases gradually with time.

iii. Activity Window (AW)-F3:

Figure 4(c) shows the total number of active days of a reviewer. 80% of spammers showed a short activity of around 3 months on the system. On the other hand, almost 50% of non-spammers have shown activity on the system for more than 12 months and the rest 50% have an activity for less than 12 months.

iv. Review Count (RC)-F4:

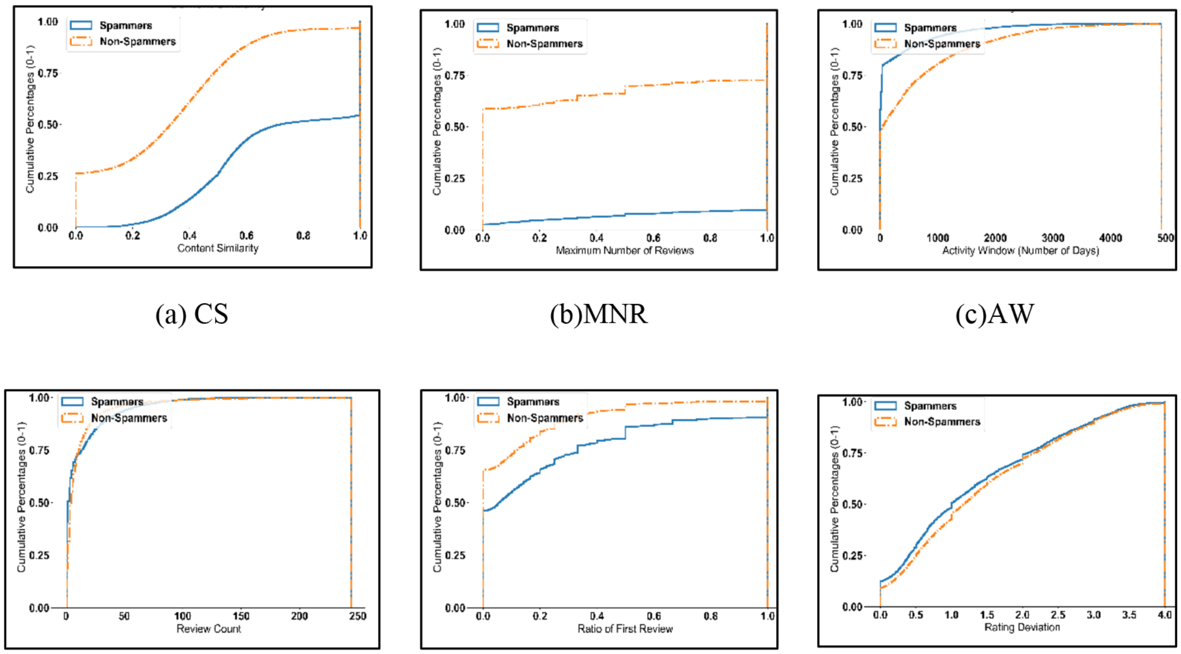
Because of short activity, spammers also have a lesser number of total reviews ever posted by them on any website. Figure 4(d) shows that approximately 75% of spammers are bounded by not more than 25 reviews throughout the activity. On the other hand, approximately 60% of non-spammers have less than 10 reviews, whereas the rest 40% have more than 10 reviews up to a total of maximum of 240 reviews as per our dataset.

v. The ratio of First Review (FR)-F5:

Spammers usually prefer to review early on products to have a greater influence on other buyers. Figure 4(e) shows that around 55% of spammers have a high ratio of writing the first review on any specific product. For non-spammers, 65% have never written any first review on products and rest 35% have shown moderate behaviour in writing first reviews.

vi. Rating Deviation (RD)-F6:

It can be observed from Figure 4(f) that spammers deviate their ratings more rapidly than that of non-spammers. Around 40% of spammers are bound by rating deviation of 1 and rest 60% have more than 1 rating deviation. For non-spammers, their behaviour is quite the same, but their graph is less steep.



(a) CS (b)MNR (c)AW

(d) RC (e)FR (f)RD

Figure 4: CDF of Reviewers (Spammer and Non-Spammer) Behavioural Features.

6. Conclusion

Online reviews play an important role in many e-commerce applications because these applications depend upon user opinions and their experiences. Some spammers apply opinion spam to deceive others. So, research on spam review detection has been quite in the focus for last few years. In this work, researchers used six different spammer behavioral features for spotting the spam reviews. Next, researchers studied the weight-based method to find spam reviews. The experimental results show that the four new introduced features play an important role in spam review detection. The results of experiments show that the weight-based method can detect spam reviews effectively because different threshold values are used to identify the spam reviews. Future research should also focus on the availability of standard labelled datasets for the researcher to train the classifier, and more attributes should be added to the dataset to improve the accuracy and reliability of the spam review detection models, such as IP address of the spammer and location where the reviewer has been signed in to write the review. Furthermore, there is a need for in-depth research on the detection of spam in multilingual reviews.

7. References

Hussain, N., Turab Mirza, H., Rasool, G., Hussain, I., & Kaleem, M. (2019). Spam Review Detection Techniques: A Systematic Literature Review. *Applied Sciences*, *9*(5), 987.

Bajaj, S., Garg, N., & Singh, S. K. (2017). A Novel User-based Spam Review Detection. Procedia Computer Science, 122, (pp. 1009-1015).

Biradar, J. G. (2017). The exponential Distribution model for Review Spam Detection. International Journal of Advanced Research in Computer Science, 8(3), pp. 938-947.

Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. S. (2013). What yelp fake review filter might be doing? In: International Conference on Web and Social Media (pp. 409-418).

Fusilier, D. H., Montes-y-Gómez, M., Rosso, P., & Cabrera, R. G. (2015). Detecting positive and negative deceptive opinions using PU-learning. *Information processing & management*, *51*(4), 433-443.

Ong, T., Mannino, M., & Gregg, D. (2014). Linguistic characteristics of shill reviews. Electronic Commerce Research and Applications, 13(2), (pp.69-78).

Dematis, I., Karapistoli, E., & Vakali, A. (2018). Fake Review Detection via Exploitation of Spam Indicators and Reviewer Behavior Characteristics. In: International Conference on Current Trends in Theory and Practice of Informatics (pp. 581-595). Edizioni Della Normale, Cham.

Jindal, N., & Liu, B. (2008, February). Opinion spam and analysis. In *Proceedings of the 2008 international conference on web search and data minin*g (pp. 219-230). ACM..

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In: Mining text data (pp. 415- 463). Springer US.

Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., & Yan, X. (2018). Measuring Customer Agility from Online Reviews Using Big Data Text Analytics. Journal of Management Information Systems, 35(2), (pp. 510-539).

Chakraborty, M., Pal, S., Pramanik, R., & Chowdary, C. R. (2016). Recent developments in social spam detection and combating techniques: A survey. Information Processing & Management, 52(6), (pp. 1053-1073).

Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013). Spotting opinion spammers using behavioral footprints. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 632-640). ACM..

Heydari, A., Tavakoli, M., & Salim, N. (2016). Detection of fake opinions using time series. Expert Systems with Applications, 58, (pp. 83-92).

KC, S., & Mukherjee, A. (2016). On the temporal dynamics of opinion spamming: Case studies on Yelp. In Proceedings of the 25th International Conference on World Wide Web (pp. 369-379). International World Wide Web Conferences Steering Committee.

Li, H., Fei, G., Wang, S., Liu, B., Shao, W., Mukherjee, A., & Shao, J. (2017). Bimodal distribution and co-bursting in review spam detection. In Proceedings of the 26th International Conference on World Wide Web (pp. 1063-1072). International World Wide Web Conferences Steering Committee.

Kaghazgaran, P., Caverlee, J., & Alfifi, M. (2017, May). Behavioral Analysis of Review Fraud: Linking Malicious Crowdsourcing to Amazon and Beyond. In *Eleventh International AAAI Conference on Web and Social Media*.

Viviani, M., & Pasi, G. (2017). Quantifier guided aggregation for the veracity assessment of online reviews. *International Journal of Intelligent Systems*, *32*(5), 481-501.

Wang, Zhuo, Songmin Gu, and Xiaowei Xu. "GSLDA: LDA-based group acc detection in product reviews". Applied Intelligence 48.9 (2018): 3094-3107.

Kaghazgaran, P., Caverlee, J., & Squicciarini, A. (2018, February). Combating crowdsourced review manipulators: A neighborhood-based approach. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (pp. 306-314). ACM.

Ganesan, K., & Zhai, C. (2012). Opinion-based entity ranking. Information retrieval, 15(2), 116-150.

Hazim, M., Anuar, N. B., Ab Razak, M. F., & Abdullah, N. A. (2018). Detecting opinion spams through supervised boosting approach. PloS one, 13(6), e0198884.

Zhou, W., Liu, M., & Zhang, Y. (2017, December). Detecting Spammer Communities Using Network Structural Features. In *International Conference on Collaborative Computing: Networking, Applications and Work-sharing* (pp. 670-679). Springer, Cham

Asadi, R., Kareem, S. A., Asadi, M., & Asadi, S. (2015). A single-layer semi-supervised feed forward neural network clustering method. *Malaysian Journal of Computer Science*, *28*(3), (pp. 189-212)

Pudaruth, S., Moheeputh, S., Permessur, N., & Chamroo, A. (2018). Sentiment Analysis from Facebook Comments using Automatic Coding in NVivo 11, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 7 N. 1, (pp. 41-48).