International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Spam Reviews and Spammer Community Detection using Machine Learning Algorithms

Nishigandha Rananaware[1], Shweta More[1], Prajakta Jagtap[1], Anuja Kumbharkar[1], A. C. Jadhav[2]
[1]Department of Computer Engineering, NESGI, Pune, India
[2]Assistant Professor, Department of Computer Engineering, NESGI, Pune, India

Article Info
Volume 8, Issue 3
Page Number: 444-449

Publication Issue :
May-June-2021

Article History
Accepted : 10 June 2021
Published: 17 June 2021

ABSTRACT

Online reviews and feedback of a product plays a vital role in human tendency to purchase those products. To affect the product sale spammer generates fake reviews on online social media platform. To identify spam reviews and spammer communities is the area of interest of this research work. In literature work, various spam detection techniques are proposed based on Review-Behavioral (RB) Based features, Review-Linguistic (RL) Based Features, User-Behavioral (UB) Based Features are explained but none of the technique provide a simultaneous study of these features and weighting of the features along with finding the relationship among the spam users. The proposed work generates a hybrid feature selection method which merge linguistic based features and behavioral features along with NLP processing and sentiment analysis. Also deep learning classification is used. The results show 91 % accuracy for detecting spam reviews.

Keywords : CNN, spam reviews, machine learning, social network.

I. INTRODUCTION

The product is judged by the customer based on online reviews. These online reviews have an impact on product sales. An increasing number of firms are attempting to obtain online appreciation from their customers in order to make a positive image in the market. Negative marketing of the competitor's product, on the other hand, is promoted in order to achieve a superior position. Artificial reviews are manufactured and uploaded on various online social media sites to float positive or bad evaluations dependent on corporate objectives. According to the report [4,] one-third of customer reviews on social networking platforms are suspect. Spamming is the act of creating phoney reviews and posting them on social networking sites.

Review spam has become so prominent that Spam individuals or spammer groups are publicly soliciting their services on websites like Fiver and Facebook, with a single review costing anywhere from $5 to $15 [14]. A recent news report by Fox News said that multiloads of Amazon Sellers are manipulating reviews on their own products to increase sales and gain product popularity. Even though websites like Amazon, Yelp and Dianping have strict Spam filters in place, most of these Spam filters can be further

improved to accommodate the complex nature of spam, that allows them to pass through these filters undetected. This paper critically analyses the existing studies, methodologies, techniques, and algorithms applied for opinion spam detection. Furthermore, the purpose of this work is twofold: To detect spam reviews using machine learning algorithms and to used various features extraction techniques for such as NLP techniques and The three methodologies i.e Review-Behavioral (RB) based features, Review-Linguistic (RL) based features, and User-Behavioral (UB) based features.

II. LITERATURE WORK

Various spam detection algorithms are discussed in the literature. These approaches are primarily divided into three groups:

2.1 Linguistic Based

Natural language processing techniques are utilised in the linguistic method to identify similarity among numerous reviews. The n-gram and its composition are used by Feng et al. [3]. Language modelling is also used in some studies [2, 4] to look for similarities between numerous evaluations, such as capital terms in assertions. To detect similarities between many reviews, Lai et al. [5] proposes the probabilistic language modelling technique.

2.2 Behavioral Based

This method is based on a review's metadata analysis. The analysis of user behaviour and review behaviour is included in the metadata. Feng et al. [6] present a method for analysing review information based on the distribution of user ratings on various products. Jindal et al. [7] provide 36 alternative behaviour analysis strategies using supervised learning mechanisms [11]. Behavioral features are more effective than linguistic features in identifying spammers. To identify spam reviews, Fi et al. [12] proposes a machine learning algorithm. Syntactic stylometry for deception detection is investigated in paper [13].

2.3 Graph Based

Spam detection can be done using network-based techniques. This strategy creates a heterogeneous network between reviews and users. In [8], Fei et al. introduced a network-based Loopy Belief Propagation (LBP) technique for detecting burstiness in reviews in order to detect spam reviews. In [10], Li et al. offer a method for analysing reviews from many users with the same IP address. Users, reviews, and user IPs form a heterogeneous network for this purpose.

Each category is investigated independently. Saeedreza Shehnepoor et al proposed the Netsapm[1] approach. This technique proposes a simultaneous investigation of behavioural (RB), linguistic (RL), and graph-based approaches. In review systems, EuijinChoo, Ting Yu, and Min Chi [9] detect spammer groups. This is accomplished through the application of graph theory and sentiment analysis on user interactions. It examines the user relationship graph, annotates it using sentiment analysis, and then prunes it. According to research in the literature, a common platform is required for the study of spam reviews and the link between various spam detection systems, as well as the identification of spammer communities.

III. PROPOSED SYSTEM

A. System Architecture:

Figure 1 depicts the stages involved in locating spam reviews and spammer communities:

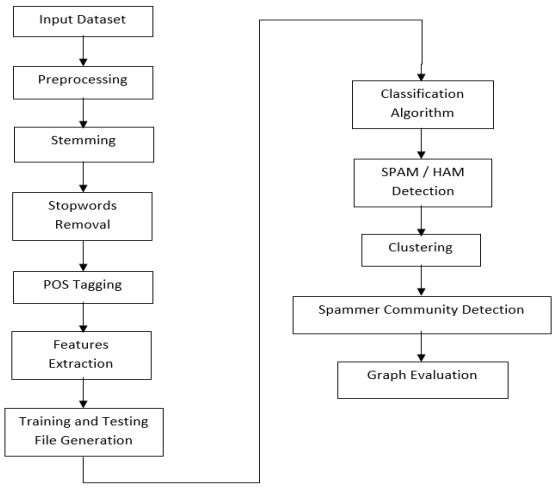

Figure 1: System Architecture

When data from a given training dataset is received, the first step is to perform preprocessing and then stemming and stopwords removal process is performed to get dictionary words after this NLP pos tagging is performed and NLP features are extracted from reviews. RL, RB features are also extracted and then training and testing file are generated which are given to the classification algorithm and at the end graph evaluation is performed to compare the results of algorithms.

B. Algorithms

1. Proposed System Algorithm

Step 1: Input Dataset:
Step 2: Tokenization
Step 3: Stemming
Step 4: Stopwords Removal
Step 5: POS Tagging
Step 6: Feature Extraction
Step 7: Machine Learning Classification (SVM)
Step 8: Deep Learning Classification
Step 9: Clustering (Spammer Community Detection)

2. Classification Algorithm (CNN Algorithm):

We're going to build a CNN with the following architecture:

1. Input Layer
2. Convolutional Layer
3. ReLU Layer
4. Pooling Layer
5. Softmax Layer

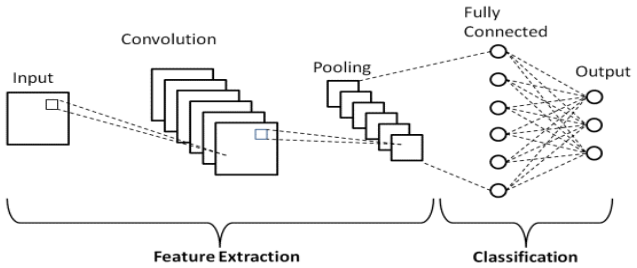

Figure 2: CNN Architecture

Step 1: Input Layer

Sequence x contains n number of entries. Every entry is represented by a d-dimensional dense vector; in this way the input x is represented as a feature map of dimensionality d {n}.

For an input sequence with n entries: x1; x2: xn, where n be total the number of features in dataset and x be the feature.

Step 2: Convolution Operation

Our initial strategy is to use a convolution process. We'll focus on feature detectors in this section, which are effectively the neural network's filters.

$$(f * g)(t) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau$$

The yields of a linear operation for example, convolution are then passed through a nonlinear activation function. The most widely recognized nonlinear activation function used presently is the rectified linear unit (ReLU), which performs the following function:

$$f(x) = \max(0, x)$$

Step 3: ReLU Layer

Later, the Rectified Linear Unit or ReLU is involved. It is the most generally used activation function. Chiefly implemented in hidden layers of Neural network.

- **Equation: - A(x) = max (0, x)**. It gives output x if x is positive and 0 otherwise.
- **Value Range: -** [0, inf)
- **Uses: -** ReLu is less computationally costly than sigmoid and tanh because it includes less difficult numerical tasks. At a time only a few neurons are activated making the network sparse productive and simple for calculation.

Step 4: Pooling

The pooling operation entails sliding a two-dimensional filter over each channel of the feature map and aggregating the features within the filter's coverage zone. The dimensions of output received after a pooling layer for a feature map with dimensions nh x nw x nc are

$$(n_h - f + 1) / s \times (n_w - f + 1)/s \times n_c$$

where,

n_h - height of feature map

n_w - width of feature map

n_c - number of channels in the feature map

f - size of filter

s - stride length

A typical CNN model architecture consists of a series of convolution and pooling layers placed one on top of the other.

Step 5: Full Connection

At this stage everything that we covered all through the section will be merged together. The two processes described before i.e.: convolutions and pooling, can been thought of as a feature extractor, then we pass this features, usually as a reshaped vector of one row, further to the network, for instance, a multi-layer perceptron to be trained for classification.

Step 6: SoftMax

The softmax function is likewise a type of sigmoid function yet is convenient when we are trying to handle classification problems. The standard (unit) softmax function defined by the formula;

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

- **Uses: -** It is used when there are more than two classes. The softmax function would squeeze the outputs for each class somewhere in the range of 0 and 1 and would also divide by the sum of the outputs.
- **Output: -** It is used in the output layer of the classifier where we are actually trying to attain the probabilities to characterize the class of each input.

3. Bisect K-means Algorithm

1. Initialize the list of clusters to accommodate the cluster consisting of all points.
2. repeat
3. Discard a cluster from the list of clusters.
   {Perform several "trial" bisections of the selected cluster.}
4. for *i* = 1 to *number of trials* do
5. Bisect the selected clusters using basic K-means.
6. end for
7. Select the 2 clusters from the bisection with the least total SSE.
8. until Until the list of clusters contain *'K'* clusters

IV. RESULT AND ANALYSIS

A. Datasets:

Datasets used in the proposed work are from Amazon website. The Amazon dataset is used to evaluate the work on unsupervised mode. On request, the datasets are downloaded. The Amazon dataset is download from the link http://jmcauley.ucsd.edu/data/amazon.

B. Evaluation Parameters

Precision(P):

P = (TP / (TP + FP))

Recall(R):

R = (TP / (TP + FN))

F1-Score (F1):

F1 = (2* ((Precision * Recall)/(Precision + Recall))

Accuracy(A):

A = ((TP + TN) / (TP + FP + FN + TN))

Where TP denotes True Positive

FP denotes False Positive

TN denotes True Negative

FN denotes False Negative

C. Graph Comparison

Accuracy comparison graph showing that the which algorithm given highest accuracy among the all. Here in below screenprint the CNN gives the more accuracy than the SVM. The comparison of accuracy of the three algorithms is tabulated in table 1 and graphically represented in figure 3.

Table 1: Accuracy Comparison of Algorithms

Algorithm	Accuracy
Support Vector Machine	82
Convolution Neural Network	91

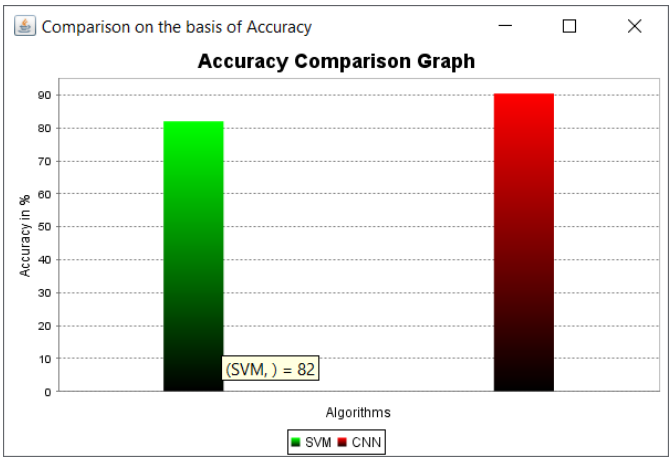

Figure 3: Accuracy Comparison Graph

Time comparison graph is used to show the which algorithm takes how much time to execute. Here in the printscreen depict CNN takes the more time to execute rather than the SVM. The algorithm execution time comparison of the SVM and CNN is graphically represented in figure 4.

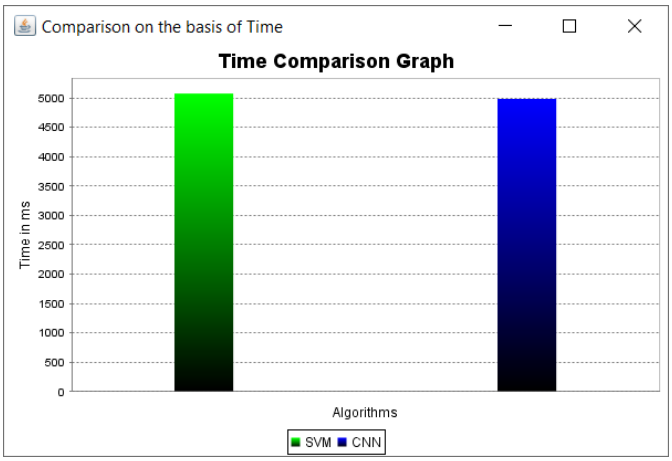

Figure 4 : Time Comparison Graph

V. CONCLUSION

The proposed system is designed for detection of spam reviews and spammer communities using Heterogeneous Information Network, which is based on metapath concept and graph based methods to label reviews. Four types of features will be used for detection of spam as Review-Linguistic Based features,

Review-Behavioral Based features, User-Behavioral Based features and User-Linguistic Based Features. We can determine the importance of each feature in classification of review and calculate the weight of each review. We can also label each review by calculating the probability of each review being spam or not. As a contribution, the spammer communities can also be detected by the system.

VI. REFERENCES

[1]. Saeedreza Shehnepoor, Mostafa Salehi*, Reza Farahbakhsh, Noel Crespi NetSpam:a Network-based Spam Detection Framework for Reviews in Online Social Media IEEE Transactions on Information Forensics and Security 2017.

[2]. J. Donfro, A whopping 20 percent of yelp reviews are fake.

[3]. M. Ott, C. Cardie, and J. T. Hancock. Estimating the prevalence of deception in online review communities. In ACM WWW, 2012.

[4]. M. Ott, Y. Choi, C. Cardie, and J. T. Hancock. Finding deceptive opinion spam by any stretch of the imagination.In ACL, 2011.

[5]. Ch. Xu and J. Zhang. Combating product review spam campaigns via multiple heterogeneous pairwise features. SIAM International Conference on Data Mining, 2014.

[6]. N. Jindal and B. Liu. Opinion spam and analysis. In WSDM, 2008.

[7]. F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[8]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh. Exploiting burstiness in reviews for review spammer detection. In ICWSM, 2013.

[9]. Choo E., Yu T., Chi M. (2015) Detecting Opinion Spammer Groups Through Community Discovery and Sentiment Analysis. In: Samarati P. (eds) Data and Applications Security and Privacy XXIX. DBSec 2015. Lecture Notes in Computer Science, vol 9149. Springer, Cham.

[10]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao. Spotting fake reviews via collective PU learning. In ICDM, 2014.

[11]. N. Jindal, B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. In ACM CIKM, 2012.

[12]. F. Li, M. Huang, Y. Yang, and X. Zhu. Learning to identify review spam. Proceedings of the 22nd International Joint Conference on Artificial Intelligence; IJCAI, 2011.

[13]. S. Feng, R. Banerjee and Y. Choi. Syntactic stylometry for deception detection. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers; ACL, 2012.

[14]. Jillian D'Onfro, ―Google Maps spam fighters,‖ CNBC, 2018. Online]. Available: https://www.cnbc.com/2018/04/13/googlemaps-spam-fighters.html. Accessed: 24-May-2018]

Cite this article as :

Nishigandha Rananaware, Shweta More, Prajakta Jagtap, Anuja Kumbharkar, A. C. Jadhav, "Spam Reviews and Spammer Community Detection using Machine Learning Algorithms", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 8 Issue 3, pp. 444-449, May-June 2021.

Journal URL : https://ijsrset.com/IJSRSET2183197