Spam Review Detection using Linguistic Method By Deep Learning Classifiers - A Survey

Atharva Narkhede[1], Prajwal Satpute[1], Priyanka Valesha[1], Srushti Pawar[1], Prof. Kshama Balbudhe[1]
[1]Department of Information Technology, Pune Vidhyarthi Griha's College of Engineering & Technology, Pune,India
Email : narkhedeatharva00@gmail.com, prajwal.satpute2000@gmail.com, priyankavalesha14@gmail.com, srushti.pawar112001@gmail.com, ksb_it@pvgcoet.ac.in

***Abstract*— Online shopping is the new fashion among the newer generation of the people. People of this generation are keen to shop various things from the online shopping sites as it is easy to purchase products from home. The reviews of those products are very impactful for the selling of the product. The manufacturers are also extremely concerned with various online reviews. Unfortunately, to increase gain and profit, the amount of spam reviews have increased in recent times. In recent years, the spam review detection problem has gained much attention from communities and researchers. In this work, we have used the amazon review dataset. As the amazon dataset is unlabelled so we have used the various behavioral features to label the dataset as spam and not spam. Spam review detection using linguistic method is used on content of review and utilizes the various transformations, feature selection and classification to identify spam review. The three deep learning classifiers are Convolutional Neural Network(CNN), Bi Directional RNN, Multilayer Perceptron(MLP).**

***Index Terms*— Online product reviews, spam review detection, linguistic features.**

I. Introduction

The use of the world wide web(WWW) has increased in recent times. As all the people have access to the internet, the trend of online shopping is increasing each day. People are easily writing blogs, forums and sharing their opinion about various products. Many of the people are considering the online opinions while buying anything from the e-commerce sites[1]. For Example, if anyone is going to buy the mobile phone, he will definitely check the various reviews to check the functioning of the product, issues related to design, information about the manufacturer. Recently the trend of spam reviews is increasing as anyone can write the review [2]. The spam review is the review which is written to promote or demote the product. The main problem about spam review is there will be false hype about any product which inturn easily distracts the public from buying a genuine product.

In recent years, many public opinion sharing websites have started some progress in spam review detection. But, there is a lot of room for improvement in study in spam review detection techniques[3]. Spam reviews give the wrong information about the product and are very much difficult to detect manually. Spam review acts like legitimate until we apply the various behavioral features on it. Based on this perspective, our proposed approach uses the various behavioral features to label the dataset as spam or not spam by considering the various features of the reviewer as the time stamp, rating, etc.

From the literature survey, we learnt that the previous studies either focus on behavioral features or linguistic features separately. But, our proposed study combines both of the studies. Most of the existing studies have used the uni-gram linguistic approach but in our study we have also used the bi-gram and trigram approach as it gives better results. In this study, the investigation is carried out on the amazon review dataset. However, the main limitation is the dataset is unlabelled. To tackle this problem we have used the Behavioral features to label the dataset as spam and not spam. The labeled dataset, then, utilizes the linguistic features and trains using the deep learning classifiers. Specifically, the proposed study uses the various spammer behavioral features and linguistic features including n-gram techniques for developing spam review detection models.

II. Literature Survey

Spam review detection using the spammer behavioral method finds the unusual spammer patterns and relationships between different spammers. Only a few studies have explored spam review detection using the spammer behavioral method to date.For example, Mukherjee et al.[4] developed a spam review detection method using a clustering technique

by modeling the spamicity of the reviewer to identify spammer and not-spammer clusters. Heydari et al.[5] have proposed a model incorporating only time series features of the reviewer on an Amazon real dataset.
The spam review detection was first studied by Jindal and Liu in 2007 [6]. They analyzed 5.8 M reviews and found many duplicate values. They train the model using a Logistic regression classifier. Li et al.[7] have suggested the spamicity based on features such as review posting rate and temporal pattern. The dataset was from Chinese website Dianping to train the model. Ahmed and Danti [8] used various rule-based machine learning algorithms.
Lau et al.[9] have used semantic language models to identify spam reviews and train the model using the SVM classifier. Hazim et al. used features for Extreme Gradient Model and Generalized Boosted Regression Model to evaluate multilingual dataset. They got better results for both the models. Kumar et al.[10] have proposed the supervised learning model. This method analyzed the behavioral features and their interaction using multivariate distribution. Li et al.[11] used the feature based sparse additive generative model and SVM classifier to train the model. Lin et al.[12] proposed various time sensitive features to find spam reviews and train the model using the SVM classifier.
Based on literature surveys, most of the existing studies do not use the behavioral features and the linguistic features together and use only one classifier to train the model. In our work we have used the three deep learning classifiers to train the model which inturn will help us in getting better overall accuracy.

III. Dataset Review

One of the major challenges in spam review detection is to collect the labeled dataset as supervised learning uses labeled dataset. This work uses the amazon review dataset which is the real time dataset publicly available on the internet. The dataset is available in various categories. For the labeling purpose we have selected the six categories which contain 26.7 Millions of reviews by 15.4 Millions of reviewers of 3.1 Millions of products. The linguistic approach needs to have the labeled dataset. As the amazon review dataset is an unlabeled dataset, we have used the various behavioral features to label the dataset.

IV. Proposed System

A. *Labeling of dataset using spammer behavioral features :*

As the amazon dataset is an unlabelled dataset, our study uses the various spammer behavioral features to label the dataset. The output of this section is the labeled dataset having the labeled as spam and not spam. The labeling of the dataset is carried out in 4 phases.

A. *Spammer Behavioral Feature :*

The labeling is done using the 13 behavioral features. Firstly, the normalized value for each of the 13 features is calculated. The 13 behavioral features used to label the amazon review dataset are as follows.

A. Content Similarity - F1: As most of the spammers usually copy the content of the already available reviews, Content similarity is one of the most important features in behavioral features. The cosine similarity function is used to find the content similarity between the review text.

B. Maximum Number of reviews - F2: Many of the spammers write many reviews in a single day. This is abnormal behavior as compared to the genuine reviewer. So, we have found the ratio of the most number of the reviews by the reviewer in a single day to the total number of reviews by the reviewer altogether.

C. Review Brustiness - F3: As the spammers write a large number of the reviews in a short time, we can consider this behavior as unusual. Our proposed study calculates the total number of reviews on a single day. If the count exceeds a certain threshold value the review is more likely to be spam. By experimental analysis, the threshold value is selected as 12. If the count of reviews is more than 12 then the normalized value of this feature is 1, otherwise 0.

D. Activity Window - F4: As the spammer is not a long time member of the website, Our study calculates the difference between the timestamp of the first review and the last review. If the difference between the first and last review is less than the 45 days then the normalized value 1 is assigned otherwise 0.

E. Review count - F5: As the spammer is not a long time member of the website, the number of reviews by spammers are less than that of the reviews by the genuine reviewers. Here, we have taken the threshold value as 5, if the review count by the reviewer is less than 5, the normalized value 1 is assigned, otherwise 0 is assigned.

F. The ratio of positive reviews - F6: The proposed study calculates the number of positive reviews i.e the review with rating 4 & 5 by the particular reviewer. This suggests that the particular reviewer is inclined towards writing the positive reviews. It is the ratio of the number of reviews with rating 4 & 5 to the total number of reviews by the reviewer.

G. *The ratio of negative reviews - F7:* Just like positive reviews, the negative reviews are also important while labeling the reviews. The proposed study calculates the number of negative reviews i.e the review with rating 1 & 2. It is the ratio of the number of the reviews with rating 1 & 2 to the total number of reviews by the particular reviewer.

H. *The ratio of first review - F8:* Our proposed approach calculates the total number of first reviews to the total number of reviews by the particular reviewer. The spammer usually writes the first review of the product to create more impact on the buyer. So, this behavioral feature is also an important one.

I. *Review of single product - F9:* Usually, the spammers are hired to write reviews of any particular product. Using this feature, our study detects the reviewers which write the reviews for a single product are marked as spam.

J. *Rating deviation - F10:* Usually, the spammers are used to promote or demote the product. So, the spammers give the rating which is different from the overall mean rating of the product. Using the feature, the mean rating for the product is calculated. Based on the mean rating value of the product the normalized value is assigned.

K. *Review length - F11:* Usually, the length of the review by the spammer is less than the review length of the genuine reviewer. Here, we have used the threshold value for review length as 400. If the length of the review is more than 400 characters then the normalized value assigned as 0 else 1.

L. *Extreme rating - F12:* Usually, the spammer gives the extreme rating i.e the rating of 1 & 5 as their main aim is to falsely promote and demote the product. If the rating is extreme than the normalized value of 1 is assigned else 0.

M. *The ratio of Capital letters - F13:* Usually, spammers write the reviews in capital letters as the capital letters are more eye catcher. Our proposed approach calculates the ratio of total count of capital letters to the total number of characters in the review.

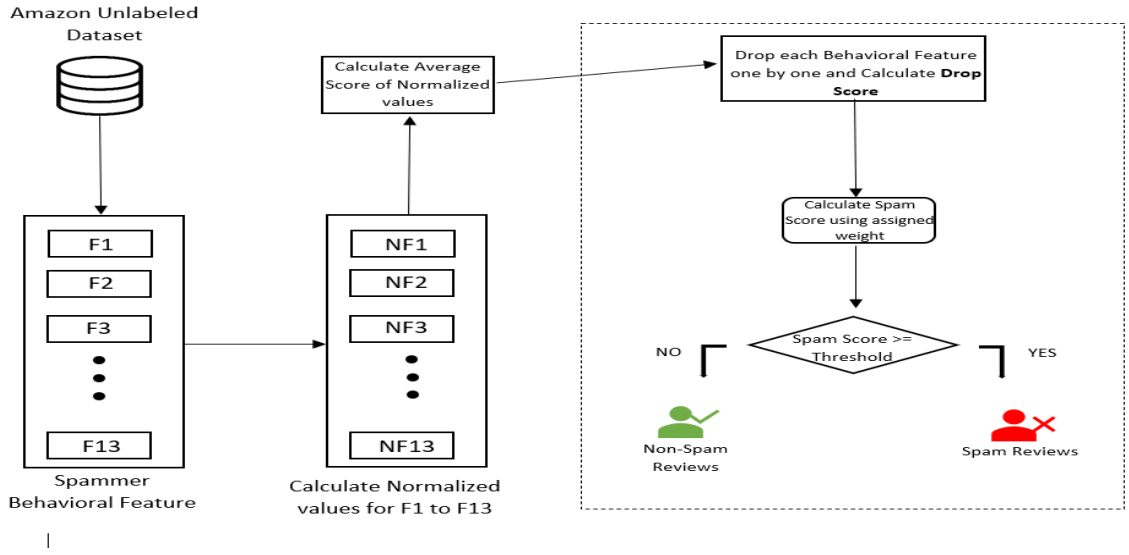


Figure 1 : Labeling of reviews

B. *Mean value of all normalized value :*

In this phase, the mean of all the normalized values i.e the values which are generated by the using the behavioral features is calculated.

C. *Dropping individual spammer behavioral feature method :*

If the mean of the normalized value is dropped by 5 % or more than the weight value of "2" is assigned, else the weight value of "1" is assigned. After assigning the weight to individual features the total weight is calculated by adding all the individual weights.

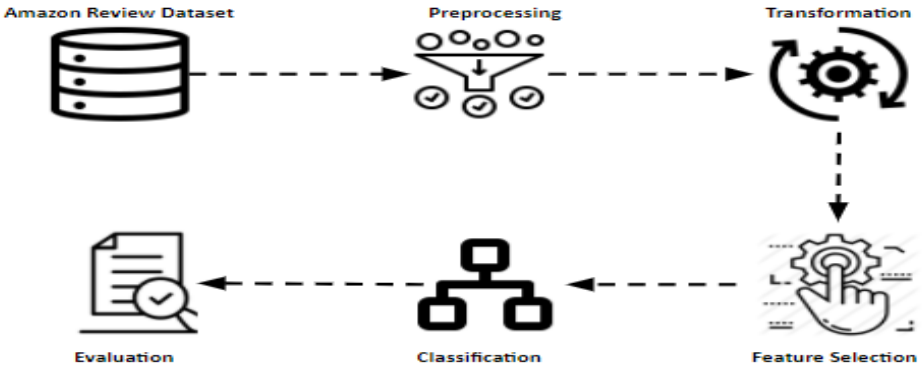
D. *Spam score method :*

In this phase, the final spam score is calculated which is used to label the review as spam and not spam. In the process, the score for each feature is calculated by multiplying the normalized value and the weight assigned to that feature. All the score values for the feature are then added to find the total score value. The spam score is the ratio of the total score

value to total weight value which is calculated in the previous phase. After calculating the spam score, the score is compared to the threshold value of 0.5. This threshold value is calculated by the experimental analysis. If the spam score is more than the threshold value then the review is labeled as Spam and if the spam score is less than that of the threshold value the review is labeled as Not spam.

B. *Spam Review Detection Using Linguistic Method :*

The linguistic method uses various pre-processing techniques, transformations and feature selection and various deep learning classifiers. This process is carried out in 6 steps.



SPAM REVIEW DETECTION USING THE LINGUISTIC METHOD

A. Dataset

The labeled dataset obtained from the previous steps of labeling is used as the input dataset in this step. This dataset is used for the training and testing of the Spam review detection model.

B. *Pre-Processing*

The pre-processing of dataset includes

Removing Stop Words & Punctuation : The review text contains some of the unnecessary words like 'a', 'an', 'the', 'is', 'are' and more. These words are not useful in spam review detection. These words and punctuations only increase the unnecessary noise in the dataset. So, these unnecessary tokens are removed from the review text.

Stemming Word : Stemming is used to convert the different forms of the words into similar form. For Example, if the text contains the words like 'works', 'working', 'worded', etc. These words are converted into the simple word 'work'. This increases the accuracy of the dataset.

Tokenizing : This technique is used to split the text into a sequence of words. Here, we have used the N-gram tokenizing technique. While tokenizing the uni-gram combination contains a single word, bi-gram contains two words while Tri-gram combination contains 3 words.

C. *Transformation*

As we have are using the textual data and the linguistic method uses numeric data. The textual data is converted into the numeric data by the help of the matrix form. Here, we have formed the sparse matrix. The sparse matrix contains the frequency of tokens in review. Here, the TF-IDF has been applied to transform the text review into numerical form.

D. *Feature Selection*

Feature selection technique will be used to select the most appropriate features which appear in the review dataset.

E. *Classification*

After converting the text reviews in the matrix format those matrices are supplied to the three deep learning classifiers for the classification.

Convolutional Neural Network (CNN): Convolutional Neural Networks are the class of the artificial neural network(ANN). They are regularized versions of multilayer perceptrons. CNNs take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filter. Also, the preprocessing needed to CovNet is very less than other DL classifiers.

Bidirectional Recurrent Neural Network (Bi - RNN): Bidirectional LSTM are the extension of the traditional LSTMs. They are used to improve the model performance. Bidirectional LSTM trains two instead of

one LSTM on input sequence. First on the input sequence as it is and second on the reversed copy of the input sequence. This provides additional context to the network and results in faster and filler learning on the problem.

Multilayer Perceptron (MLP): The MLP is the class of the feedforward artificial neural network(ANN). It consists of at least three layers: an input layer, a hidden layer & an output layer. Each node is a neuron that uses a non-linear activation function. MLP utilizes a supervised learning technique called backpropagation for training. MLP can distinguish data that is not linearly separable.

F. *Evaluation*

As the classification is done, the evaluation is carried out by considering various parameters like recall, precision, f-score, accuracy.

V. Conclusion

Review spamming is a rapidly growing problem in the case of online shopping sites. This problem should be tackled by us if we want to ensure the customer is purchasing the right product. But, Spam review detection is a very difficult task to state whether the review is spam or not. Our work will differentiate the reviews as spam and not spam using 13 behavioral features. The features like content similarity, review of single product, activity window are very useful in spam review detection. This feature will be promptly used to label the dataset. The linguistic method will definitely improve the accuracy and reliability of the work. The deep learning classifiers will definitely perform better than the traditional machine learning classifiers. They will be useful to increase the reliability of the work.

REFERENCES

[1] J. Huang, T. Qian, G. He, M. Zhong, and Q. Peng, ‘‘Detecting professional spam reviewers,’’ in Proc. Int. Conf. Adv. Data Mining Appl. Berlin, Germany: Springer, 2013, pp. 288–299.

[2] S. Bajaj, N. Garg, and S. K. Singh, ‘‘A novel user-based spam review detection,’’ Procedia Comput. Sci., vol. 122, pp. 1009–1015, Jan. 2017.

[3] J. G. Biradar, S. P. Algur, and N. H. Ayachit ‘‘Exponential distribution model for review spam detection,’’ Int. J. Adv. Res. Comput. Sci., vol. 8, no. 3, pp. 938–947, 2017.

[4] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, ‘‘Spotting opinion spammers using behavioral footprints,’’ in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2013, pp. 632–640.

[5] A. Heydari, M. Tavakoli, and N. Salim, ‘‘Detection of fake opinions using time series,’’ Expert Syst. Appl., vol. 58, pp. 83–92, Oct. 2016

[6] N. Jindal and B. Liu, ‘‘Analyzing and detecting review spam,’’ in Proc. 7th IEEE Int. Conf. Data Mining (ICDM), Oct. 2007, pp. 547–552.

[7] H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee, and J. Shao, ‘‘Bimodal distribution and co-bursting in review spam detection,’’ in Proc. 26th Int. Conf. World Wide Web (WWW), 2017, pp. 1063–1072.

[8] S. Ahmed and A. Danti, ‘‘Effective sentimental analysis and opinion mining of Web reviews using rule-based classifiers,’’ in Computational Intelligence in Data Mining, vol. 1. New Delhi, India: Springer, 2016, pp. 171–179.

[9] R. Y. Lau, S. Y. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, ‘‘Text mining and probabilistic language modeling for online review spam detection,’’ ACM Trans. Manage. Inf. Syst., vol. 2, no. 4, pp. 1–30, 2011.

[10] N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, ‘‘Detecting review manipulation on online platforms with hierarchical supervised learning,’’ J. Manage. Inf. Syst., vol. 35, no. 1, pp. 350–380, Jan. 2018.

[11] J. Li, M. Ott, C. Cardie, and E. Hovy, ‘‘Towards a general rule for identifying deceptive opinion spam,’’ in Proc. 52nd Annual. Meeting Assoc. Comput. Linguistics, vol. 1, 2014, pp. 1566–1576

[12] Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, and A. Zhou, ‘‘Towards online anti-opinion spam: Spotting fake reviews from the review sequence,’’ in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2014, pp. 261–264.