Advances and Applications in Mathematical Sciences
Volume 21, Issue 5, March 2022, Pages 2695-2705
© 2022 Mili Publications, India

A STUDY ON SPAM DETECTION WITH SENTIMENT ANALYSIS

JASNEET KAUR

Department of Information Technology
Inderprastha Engineering College
Ghaziabad (201010), Uttar Pradesh, India

Abstract

Considering the popularity of websites like Amazon, Flipkart, TripAdvisor or any social media platform, posting online reviews is a very common way to share experiences. Most of the consumer refers to these reviews before purchasing any product or service. Most people prefer buying a product having maximum positive feedbacks or 5 star rating. However, not all the reviews/comments posted online are genuine. Because of increasing market competitions, many companies are engaging spammers to publicise their products or defame the similar products of their competitors. Apart from e-commerce websites, the spammers also spread fake news or links through blogging websites, emails and SMS just to delude customers and influence their ideas.

This paper is intended to discuss research works which are conducted in the field of spam detection by various scholars and give its comparative view of the various techniques used in recent study.

1. Introduction

Sentiment Analysis, which is also known as Opinion Mining is a branch of machine learning. It involves identifying the emotion of a writer by performing various classification techniques and identifying a positive, negative or neutral feeling of a review. Review classification can be performed on any dataset involving user feedback about any article.

With an emerging role of World Wide Web in our lives, most of us prefer to search almost everything online. Whether it is something related to personal

2020 Mathematics Subject Classification: 68T50.
Keywords: Sentiment analysis, Opinion mining, Spam detection.

*Corresponding author; E-mail: jasneet.kaur@ipec.org.in; jasneet1205@gmail.com

Received September 8, 2021; Accepted November 30, 2021

usage or household, reading about an article or searching about what's going around us, we prefer to read online. This has led to a marketing strategy of influencing readers/customers through reviews. Reading the reviews before purchasing a product has become a habit of any potential customers. But few companies exploit this feature and try to increase their product sale by posting fake reviews for their products. They also hire spammers to post positive reviews on their products and negative reviews on similar products of their market competitors.

Spam, basically means any type of communication created from either a person or a group which is intended to mislead its reader by slandering some other person or entity. It can even contains unsolicited advertisements or even potentially harmful contents such as virus or malware. Some common types of spams are:

1. Spams through E-mails. These are unsolicited messages sent via E-mail. Most of the time they are sent in bulk. A Spam message can be a type of an annoying advertisement or any can also be some harmful external links which might lead to phishing websites that can steal your personal information or can contain virus or malware.

2. Product promotions. These are unwanted SMS or Emails sent by companies just to promote sale of their products

3. Citations Spams. It involves the process of using or making citations in the improper or Illegal manner is termed as a citation spam. These spams generally originate in the fields of academic articles by scholars and scientists.

4. Spam through External Link. Many companies promote some product by creating external links through some social media webpages to advertise their products.

5. Product Review Spams. These spams take place on e-commerce websites or even on websites where customer share their feedback regarding any product or entity. Most of the people tends to refer to user reviews before purchasing any product. To make use of this marketing strategy, companies hire spammers or some group of persons who post misleading reviews just to promote their sale and to deprecate similar products of competitors.

With increase in online spamming, the requirement of spam detection is becoming more important so that the customers cannot be misled by spammers. Various research scholars have proposed techniques to detect these spam. We will now have an overview of all the latest work done in this area and have a comparative view of the same.

2. Related Work

"Simran, Niharika and Sandeep have worked on the IP address of the device of a user and geographical location with which he is retrieving different resources on web. Also, they have proposed a content analysis means to detect non-reviews using spam dictionary and proposed a spam detection techniques based on four different features together." [1]

"Next work Proposed a spam review detection system which efficiently employ following three features: (i) sentiments of review and its comments, (ii) content based factor, and (iii) rating deviation. This work investigated all these features for only suspicious review list in which only those reviews were retained which received comments by peer users." [2]

"This work Proposed a technique which makes use of a public dataset which contains tweets and account information of both genuine accounts as well as spam accounts. Which is then used to make a classifier which can easily classify whether the given account is a fake account or a genuine account? After classification, sentiment analysis algorithms were applied on the tweet to find patterns among them." [3]

"A hybrid technique is proposed which uses content-based as well as graph-based features for identification of spammers on twitter platform. This technique is analysed on real Twitter dataset with almost 11k users and over400k tweets." [4]

"A hybrid set of features method is used for spam detection (Opinion Spam, Item Spam and Opinion Spammer) and also introduce a rule-based feature weighting scheme and proposed a way for tagging the review sentence as spam and non-spam." [5]

"The main objective of author was to present an enhanced feature-based sentiment analysis algorithm that improves the performance of opinion classification. The proposed algorithm is developed to assign precise sentiment score to each feature in customer's analyses in view of spam reviews detection. The proposed work also inspects the effect of three different feature extraction methods on the performance of sentiment classification."[6]

"This work undertakes the spam detection techniques using machine learning classifiers such as Logistic regression (LR), decision tree (DT) and K-nearest neighbour (K-NN) to detect ham and spam messages in mobile device communication. The SMS spam collection data set is used for testing the method. The work thus compares efficiency of various machine learning algorithms to detect which one works the best." [7]

In next work "the author mentioned that most existing methods have lower accuracy in identifying fake reviews because they just use single features and lack of categorised experimental data. To solve this problem, a method to detect fake reviews based on multiple feature fusion and rolling collaborative training. First, this method involves an initial index system with multiple features such as text features, behaviour features of critics and sentiment features of reviews. Then the method needs an initial training sample set. So the related algorithms are designed to extract all the features of a review. Finally, the method uses the initial sample set to train 7 classifiers, and the most accurate one will be selected to classify new reviews" [8]

"A new set of features are prepared by using very popular machine learning classification algorithms, namely Naive Bayesian (NB), k-Nearest Neighbor (k-NN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). The performance of these classifiers are calculated and compared on the basis of different valuation metrics" [9]

"A feature selection process is used to identify the most influencing features in the process of detecting spam twitter profiles. For feature selection, the researchers have used two methods are ReliefF and Information Gain. While for review categorisation, four classification algorithms are implemented and compared:, Decision Trees, Multilayer Perceptron, k-Nearest neighbors and Naive Bayes.The author thus concluded that some promising detection rate can be achieved using such features.” [10]

3. Comparative study

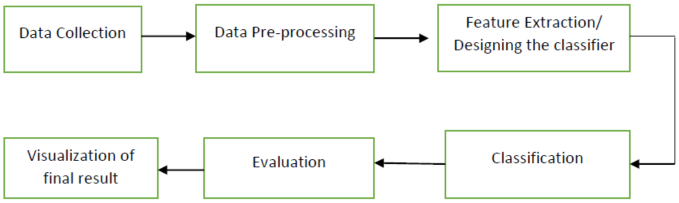
Table 1. Comparison of various studies.

Reference	Dataset Used	Methodologies/ Algorithm Used	Performance Metrics	Results
[1]	Review dataset from e-commerce website	Gmail authentication, IP Address, location and Spam dictionary	Characteristics of a reviewer were considered along with the content analysis	Unique identity of a user is maintained which is able to identify spam activity under some assumptions
[2]	Product reviews from Amazon.in	For Classification: RF,GB and SVM For Data balancing: SMOTE and ADASYN	Precision, Recall and F1 score	The proposed system achieved the F1-score of 91%.
[3]	Twitter Dataset with tweets &account information	XGBoost, Decision Tree, Random Forest, AdaBoost	Accuracy	“AdaBoost classifier with a maximum training accuracy of 99.98% and maximum testing accuracy of 92.75%.” [3]
[4]	Twitter Dataset with 11K tweets	J48, Decorate and Naive-Bayes.	Precision	“Combining user-based, content based features a significant improvement in precision for decorate

				and j48 is observed, i.e., up to 97.6%.” [4]
[5]	Amazon Based Dataset	hybrid set of features (Opinion Spam, Opinion Spammer, and Item Spam) + Rule based weighing scheme	Accuracy, Precision, Recall, and F-measure	“Revised feature weighting scheme achieved an accuracy increase from 93 to 96%. Furthermore, a hybrid set of features improve the performance of Opinion Spam detection in terms of better precision, recall, and F-measure values.” [5]
[6]	1600 reviews for Chicago-based hotels from Trip Advisor and Yelp.	Feature extraction by Apriori algorithm	Accuracy, Precision, Recall	“Proposed algorithm achieves an accuracy of around 79.56% in categorising opinions.” [6]
[7]	“SMS spam collection dataset” available on kaggle	“Logistic regression (LR), K-nearest neighbor (K-NN), and decision tree (DT)” [7]	specificity, accuracy, sensitivity, and execution time	“LR is high as compared with K-NN and DT, and the LR achieved a high accuracy of 99%.” [7]

[8]	yelp shopping website	multi-feature fusion and rolling collaborative training.	Precision, Recall, F1 score	“Accuracy of the proposed method for detecting fake reviews is 84.45%” [8]
[9]	Twitter Social Honeypot dataset	graph-based and tweet content-based features	Accuracy, Precision, Recall, F1 score	“Random Forest (RF) gives the better result compared to other ML algorithms, with an accuracy of 91%, precision 92%, and F1-score 91%.” [9]
[10]	Dataset with 82 twitter user profiles	“For feature Selection: ReliefF and Information Gain. For classification: Decision Trees, Multilayer Perceptron, k-Nearest neighbors and Naive Bayes.” [10]	Accuracy, Precision, Recall, F1 score,AUC	“Final outcome in the work show that much better results can be attained using the Naive Bayes and Decision Trees classifiers.” [10]

4. Generalised Work Flow



Figuure 1. Analysed workflow.

I. Data collection. It is the very first step of any data classification. Data is collected for spam detection from various e-commerce or social media platforms through web scrappers. Datasets which are available on web repositories like Kaggle can also be used. One can also create his own dataset using python APIs Some of the works like, [7] have used pre labelled dataset and some like [2], manually labelled the dataset, automatically or by human labelling.

II. Data Preprocessing. In this step the raw data is converted into a machine suitable form. Some of the tasks which are involved in pre-processing are:

- Data Cleaning. Data can have inappropriate and missing content So data cleaning is done to handle missing as well noisy data.
- Data Transformation. To make data suitable for classification data transformation step is applied.
- Data Reduction. To manage huge amount of data, data reduction techniques are applied. The aim is to reduce data storage cost and improve storage efficiency.

III. Feature Extraction. Feature extraction process is applied by finding nouns and noun phrases. How sentiment analysis performs relies on the efficiency of the feature extraction method used. Many works that are discussed above have used feature extraction methods in different ways. In [2], the research is done on the basis of seven extracted different features from review data and comment data. In [4], the author has mentioned about three types of features that are being used, user-based, content-based and graph-based. In, [5] a hybrid feature selection scheme is used the feature set of a baseline Spam detection method is enriches with Spam detection features (Opinion Spam, Opinion Spammer, Item spam).

In [6], three types of feature selections are used extracting all nouns, extracting only the nouns that occur frequently and extracting frequent nouns by applying Apriori algorithm.

In [8], author has used multiple feature fusion and rolling collaborative training model. In [9], For detecting Twitter spammers, author has make use of several new features, which are more effective and robust than existing used features (e.g., number of followings/followers, etc.). In [10], “The author has proposed ten simple features that can be used to classify spam profiles.” [10]

IV. Classification. The designed features are then used to develop the spam detection model. This is conducted by training different popular machine learning models. The research work discussed above use various ML models like Naive Bayesian (NB), *k*-Nearest Neighbour (k-NN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM). The models developed are then assessed and tested on a different unobserved piece of the dataset for final valuation.

V. Evaluation. The detection models created in the preceding stage are assessed using different evaluation metric. Most commonly used evaluation metric is accuracy, precision, recall and F1 measure.

VI. Results. The results of various classification/feature extraction schemes are then compared to give final result. The end results of various research works discussed above have compared various state of art techniques as well as defined their own classification rules for a better detection accuracy.

5. Conclusion and future work

Thus the major identification of this study is that there are still many issues and lack of research in detection of spam reviews, some of which are elaborated below:

I. Inaccessibility of labelled datasets. “One of the major challenges faced by researchers is lack of labelled dataset. The present datasets are either unlabelled, or they are not having adequate number of attributes required for proper training of classifiers for classifying spam and non-spam reviews” [11].

II. Rapid growing rate of review datasets. “Review-based websites, such as Yelp.com already have got hundreds of thousands of critiques which are rising unexpectedly. Such massive datasets involve unusual computing power for analysis and a major challenge is use of semantic algorithms in this field. SentiWordNet is majorly used for opinion mining which have a large repository of words that is used for analysis of reviews. Till date, a semantic-based model has not been proposed for spam review detection." [11]

III. Finite data attributes. "The available datasets of reviews have finite attributes. This drawback makes it tough for researchers to identify spam evaluations correctly. The prime challenge here is the lack of multi-dimensional datasets. To improve the precision of the algorithms more attributes are required, consisting of, Email ID of the customer/reviewer, IP address of his system and his geographical location from where he is currently logged in to the review." [11]

IV. Multilingual review spam detection. "Many times a reviewer may use language of their choice while writing a review. So far, few researchers have worked on datasets in languages other than English, such as Arabic, Chinese, or Malay. There is a need to have a detailed research study on the detection of spam in multilingual reviews." [11]

V. Analysing the review. "By considering the content of the feedback and the reviewer's behaviour to detect spam reviews, researchers have made some improvement. However, so far the reviewer's profile details has not been used by any work. Usually, there are supplement comments by other users on the given reviews. For example, many e-commerce websites ask such questions as "Did you find this review useful?" Until now, such comments on given reviews have not been used as features for spam detection." [11]

References

[1] S. Bajaj, N. Garg and S. K. Singh, A novel user-based spam review detection, Procedia Computer Science 122 (2017), 1009-1015.

[2] S. Saumya and J. P. Singh, Detection of spam reviews: a sentiment analysis approach, Csi Transactions on ICT 6(2) (2018), 137-148.

[3] G. Shetty, A. Nair, P. Vishwanath and A. Stuti, Sentiment analysis and classification on twitter spam account dataset, In 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) IEEE (2020), 111-114.

[4] M. Mateen, M. A. Iqbal, M. Aleem and M. A. Islam, A hybrid approach for spam detection for twitter, In 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (2017), 466-471.

[5] M. Z. Asghar, A. Ullah, S. Ahmad and A. Khan, Opinion spam detection framework using hybrid classification scheme, Soft computing 24(5) (2020), 3475-3498.

[6] N. M. Saeed, N. A. Helal, N. L. Badr and T. F. Gharib, The impact of spam reviews on feature-based sentiment analysis, In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (2018), 633-639.

[7] L. Guang Jun, S. Nazir, H. U. Khan and A. U. Haq, Spam detection approach for secure mobile message communication using machine learning algorithms, Security and Communication Networks, (2020).

[8] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi and X. Xiao, Fake review detection based on multiple feature fusion and rolling collaborative training, IEEE Access 8 (2020), 182625-182639.

[9] Z. Alom, B. Carminati and E. Ferrari, Detecting spam accounts on twitter. In 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2018), 1191-1198.

[10] A. Z. AlaM, J. F. Alqatawna and H. Paris, Spam profile detection in social networks based on public features, In 2017 8th International Conference on information and Communication Systems (ICICS) (2017), 130-135.

[11] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain and M. Kaleem, Spam review detection techniques: A systematic literature review, Applied Sciences 9(5) (2019), 987.