Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/6/20230734

# Deep learning for comment spam and scams

**Jiangchuan Liu**

BASIS Bilingual School, Shenzhen, China

Jiangchuan.Liu90678-bbsz@basischina.com

**Abstract**. As the popularity of online social networks has increased, so has the manner in which people purchase online. Online reviews on the purchasing of items or the provision of services have become the primary source of user opinion. When a substantial body of knowledge assesses a product or service, that body's effect on the market is significant. Because of this, concerns have been raised among manufacturers and merchants, who frequently compose rubbish reviews to promote or denigrate the quality of specific items or services in order to make a profit or maintain their reputation. You may choose to promote or denigrate certain products or services that have been targeted. This kind of commentary is known as "trash remark." Comment spam violates the interests of enterprises to a significant level, and it possesses a defensive mechanism that combines the "Deepfake" of artificial intelligence technology, which renders conventional security measures inapplicable. Deep learning is the approach that will be most effective in resolving this issue. Researchers assess this research according on how they extract features from the commentary dataset and how they extract features from the commentary dataset, employing a variety of methodologies and strategies to find solutions to the problems. The authors also looked at the key deep learning technologies that have been suggested as a solution to the issue of spam detection in the primary machine, as well as the performance of various deep learning technologies. At the same time, the fact that there is not a lot of spam online review data makes this sort of research a highly significant one. The purpose of this article is to offer an insightful and all-encompassing analysis of comparative research on the topic of spam detection that has recently been conducted.

**Keywords:** Machine learning, Deep learning Big data, Deepfake, Social spam, Spambot, Spammer.

## 1. Introduction

The influence of online reviews is increasing at a rate that is proportional to the expansion of both the size and significance of the Internet. Reading reviews to decide whether or not to purchase a product is the easiest and quickest way for users, and reading other users' reviews is the most trustworthy way from a consumer's point of view, because the very nature of social platforms is to allow people to share their lives with each other, and people tend to be more honest about the products and services they use when they share their experiences with others [1]. Reviews can affect a whole wide range of industries, but they are particularly important in the e-commerce sector, where reviews and comments on Before making a purchase, the vast majority of consumers first look at other people's opinions about the goods or service. Everyone on the internet is now coming to the realization that these online reviews are really important to other customers and vendors. Suppliers can also design their additional marketing strategies based on these reviews, which is especially helpful in light of the fact that social networking platforms and shopping platforms built specifically for shopping, both of which have seen a surge in popularity over the past few years, are increasingly turning to community templates to help customers improve their shopping experiences [2]. For instance, if multiple customers buy a specific model of laptop and write reviews about a problem, the manufacturer may realize that addressing the problem can lead to increased customer satisfaction. However, because the free-speaking nature of these platforms also gives businesses who want to monetize their products by special means the opportunity to incentivize those who write good reviews about their goods, or pay someone to write bad reviews about a competitor's products, the manufacturer may also realize that addressing the problem can lead to increased customer satisfaction [3]. Because of the significance of the evaluations, these phony reviews are referred to as spam reviews, and they have the potential to have a significant impact. The erosion of consumers' faith in a brand as a result of censorship spam can have a deleterious effect on a company's bottom line. Writing spam reviews is an extremely simple way for spammers to generate excitement about a certain product. It's possible that the worth of a product or service could be significantly increased because to these spam reviews. For instance, if a person is interested in making a purchase of a product online, they will typically head to the reviews section to read the comments and ratings left by previous customers. Users will consider purchasing the product if the majority of the reviews are positive; otherwise, they will not purchase the item. All of this demonstrates that spam reviews have become a huge issue for online buying, and that these reviews are typically massive. In certain instances, phony evaluations may even exceed actual ratings, which illustrates how widespread the problem has gotten. It is highly possible that it will result in losses for the company as well as for the customer. Typical classifications for the many types of review spam are as follows: 1. unrealistic reviews, which are the primary focus of this article; 2. regarding brands; the reviews only relate to brands or items that have not been reviewed by the seller; and 3. not reviewed; these reviews contain advertising or language that is irrelevant to the product. Experiments conducted with machine detection on a large portion of the manually retrieved fake review data revealed that Type 1 fake reviews were the most damaging type of fake review [4]. The most concerning aspect of Type 1 fake reviews was that they undermined the integrity of the online review system. Because identifying Type 1 comment spam is a difficult process, the well-known technique of "deep learning" in natural language problem solving is utilized here. This allows for the identification of more bogus comments. Deep learning, in contrast to more conventional machine learning algorithms, makes an effort to derive high-level features directly from the data itself. This is the primary distinction between deep learning and other types of machine learning. Deep learning is advantageous since it eliminates the need to manually create feature extractors for each unique challenge. Convolutional neural networks, for instance, make an attempt to learn low-level features in the layers that came before them. After that, they learn partial faces, and finally, they learn high-level descriptions of faces. You can get additional information by reading about the fascinating applications of neural network computers found inside deep learning. In the process of applying traditional machine learning algorithms to solve a problem, the problem is typically broken down into a number of smaller sub-problems, which are then solved one at a time before the final result is obtained by combining the solutions to all of the smaller problems. Traditional machine learning algorithms. Deep learning, on the other hand, encourages direct issue solving from beginning to end [5]. The primary distinction between deep learning and more conventional forms of machine learning is that the performance of deep learning improves with increasing amounts of available data. When there is a shortage of data, deep learning algorithms do not function as effectively. This is due to the fact that the deep learning algorithms require a significant amount of data in order to fully comprehend it. GPUs are primarily used to increase the efficiency of matrix operations, which is why they are required hardware for deep learning to function effectively. Since deep learning algorithms require a large number of matrix operations, GPUs are an essential piece of deep learning hardware. Deep learning requires more powerful computers with graphics processing units (GPUs) installed than the more conventional machine learning techniques do. Deep learning models incorporate domain expertise into a feature extractor to simplify the analysis of the data and produce patterns that improve the performance of the learning algorithm. This helps deep learning models minimize the complexity of the data.

Although there are already many deep learning models for detecting fake comments, the lack of data has been one of the most serious problems in this field. The results of the deep learning in this project can be used to obtain more data sets, which will help with more advanced application detection in the future.

## 2. Background

In the era of Web 3.0, entertainment through social networks has become very common, people can easily access the virtual world with portable devices, this virtual world entertainment is more important to some people than real life, many people believe in this virtual world, it is because of this characteristic this also generates huge benefits, such as advertising, because the network of This has led to fierce competition between marketers in a limited area, such as the early smear campaign between Apple and Samsung, where the competition between the two giants was fierce, and the lack of rigor in the vetting system for advertising on the internet, which led to a poor experience for users of social networks. This phenomenon has now been ostensibly solved, but the fact is that it has not, and has only moved from the surface to the shadows. The internet has evolved over time and more and more users are beginning to see the false side of the internet and trying to identify what is true and what is not, and the social networking platforms are trying to do the same, but as long as something has the potential to be profitable someone will do it, and 'spammers' have learned to pretend to be someone they can trust to deceive people [6]. This has led to a lot of effort being spent on anti-fraud efforts on social networking platforms, but this is still within the realm of control, and what really needs attention is a product of the increasingly advanced technology of the internet, "spambots "Even though these fake comments are easy to detect, they are so numerous that they sometimes exceed the number of real comments. This approach worked at first but as soon as the spammer diversified the messages sent by the spambot or even used deep learning to allow the spambot to send messages that bypassed the detection of the blocking bots, such as keywords, the name of the sender's device, IP, and so on. To cope with this diversity of false comments it is necessary to use "deep learning", through the detected and artificially provided false comments, the blocking bot can constantly change the conditions of blocking, this technology is highly evolved, and it can also clean up the web environment for the future Web 4.0 era.

For the dataset the testers chose:

Amazon Mechanical Turk, Amazon Mechanical Turk synthetic fake reviews dataset on a real-world fake reviews dataset procured from Yelp. This is an automatically generated dataset where there are a number of moderate anomalous behavioral features (percentage of positive reviews, large number of reviews, average review length, etc.), which is also very relevant to the application scenario.

Organic Solvent Nanofiltration, both graph-based and nongraph-based Organic Solvent Nanofiltration datasets are available. The graph-based dataset is the one created from social network graphs. Non-graph datasets, on the other hand, are those produced using characteristics collected from Organic Solvent Nanofiltration utilizing social network APIs. Organic Solvent Nanofiltration data is mostly collected via APIs made available by OSNs.

Different hotel reviews, these datasets are from online hotel booking website. In this dataset there are 27,952 of reviews, 1000 of Reviewers, 20,622 of Products.

Different Reviews about product, real data from Amazon, the world's largest shopping platform, in this dataset, there are 65,098 of reviews 1703 of reviewers 53,353 of products. Reviews about different restaurant，from Dianping.com，In this dataset，there are 493,982of reviews 1703 of reviewers 278 of products.

## 3. Recent Researches

Opinion spam detection framework using hybrid classification scheme [7]: The network is a huge repository of available information, allowing people from all over the world to communicate and express their feelings but the trend of writing false comments is becoming more and more serious, so opinion spam detection has become the most concentrated research field. It is reported that 15% of the comments are duplicates of old comments and may be regarded as spam. In this paper, the problem of "fake reviews to promote a target product or to discredit a competitor's specific brand" is presented and a solution is given. The authors use a dataset of reviews from amazon and sentences marked for spam detection as a dataset for testing, using spam detection features such as opinion spam, comment spam, item spam.

The authors then evaluate the role of spam-related features in detecting and classifying spam cues and distinguishing them from real reviews and introduce a rule-based feature weighting scheme and propose a method for labeling review sentences as spam and non-spam in the post-experimental the application of this technique in this area is novel and the results of this work will provide insight into the features and feature weights associated with spam. Spam detection is a relatively new and challenging field in sentiment analysis because product reviews are open, and spammers can act as different users. The results of this work will provide insight into the features and feature weights associated with spam and help develop more advanced spam detection applications.

Optimizing semantic LSTM for spam detection [8]: As the number and frequency of individuals accessing the internet increases with smartphones, it has also changed the way people express themselves and their interactions can take place on public social platforms such as Discord, Twitter. Because of the increase in traffic it also leads to users receiving spam messages they don't want to receive gaining an advantage by stealing privacy, in the report e-spamming behavior went from 87.7 percent in 2012 to 90.2 percent in 2014.This has led to strain, increased storage requirements, expansion of offensive material, such as pornographic content, and most importantly it violates the recipient's privacy.

Spam classification is a topic of ongoing research in the field of natural language, and the increase in the number of users of social networks also brings great commercial value. In this paper, the authors present the problem of "spammers trying to take commercial or non-commercial advantage by sending spam", and the solution adopted by the authors is a deep learning technique. Specially known as Long Short Term Memory, using a variant of Recurrent Neural Network, LSTM is an excellent variant of RNN model, inheriting most of the characteristics of RNN model, while solving the Vanishing Gradient problem due to gradient backpropagation process, the biggest difference between it and traditional machine classification is the features are hand-crafted.

Integrating aspect analysis and local outlier factor for intelligent review spam detection [9]: With the development of e-commerce, more and more spam reviews appear in e-commerce platforms under the inducement of profits, such as Amazon, Resellerrating, Tripadvisor. deception prevalence studies did report 8%-15% spam rate in online review sites, which seriously affects the online shopping environment, and also appeared to degrade the reputation of others' products to ensure their own sales.

In this article, the authors raise the issue of "individuals and organizations inducing people to buy products that do not match the advertised ones through online reviews". With such a large volume of information as online reviews, it is a huge challenge to intelligently detect spam reviews from the text. There are many ways to detect spam comments, such as when messages are sent and when comments are shared, but they can be detected by reducing the efficiency of sending spam comments. In this paper, the authors use an intelligent approach that can be performed automatically without supervision. The authors propose a method that generates a sentiment lexicon to compute aspect ratings of reviews and propose a local outlier model of aspect ratings to identify spam. The authors' experiments on TripAdvisor conclude that the model is effective and intelligent and that this model can greatly help the development of online web business.

A deep learning model for Twitter spam detection [10]: Social networking platforms gather a large number of Internet enthusiasts, such as Twitter is one of the popular social networking methods, discussing social issues, following favorite accounts to get news, such a huge traffic has also become the target of spam, although there have been developers developed a number of machine learning but at this stage of machine learning is still unable to effectively distinguish Twitter spam, spammer can think of bypassing the learning machine rules to allow spam to be successfully posted. The spammer can be thought to bypass the rules of the learning machine so that spam is successfully published. In the article, the authors propose the problem of "conventional machine learning cannot effectively detect spam" using deep learning to solve the problem. Deep learning can collect all known information about spammers including but not limited to, for example, the age of the account, the number of followers, the number of followers combined with the information sent to spam blocking.

The authors compare the performance of using the newest five machine learning programme and two deep learning programme, The results show that the highest performance is achieved when deep learning is used and deep learning also shows great potentials.

## 4. Challenges

The authors were only able to find a dataset that could be used in one of the hotels' reviews, and the researchers needed access to a standard labeled dataset in order to determine which reviews were spam and which were not for the purpose of classifying spam or non-spam reviews. There was a very limited amount of labelled data that could be used in the experiments, which presented a challenge for experiments that required large amounts of data.

Because of the massive amount of data that Amazon provides for model training in natural language, strong arithmetic ability is required, as well as reliance on SentiWordNet and wordnet for the analysis of a large number of sentiment words that are used. However, semantic-based spam review detection models have not been proposed as of yet.

After crawling to get the information, researchers need additional information such as the IP address of the spammer, the email address of the comment site, and the location where the commenter logged in to write the comment. Many companies that can provide large amounts of data have limited information that is not available to the public for some reason, and researchers need to crawl to get it.

Even though spammers and spammers have become more active in machine learning and deep learning since the advent of Web 2.0 and despite the fact that the techniques have shown to be effective in the last few years, spammers and spammers have become more active in machine learning and deep learning. In spite of this, they are confronted with a wide variety of difficulties, including real-time data collecting, class imbalance, the spam drift problem, feature engineering, scalability, and adversarial attacks.

The previously proposed deepfake problem, deepfake uses a large number of deep learning techniques, including but not limited to the use of GPUs, artificial intelligence, facial mapping, and neural networks; learning from large sample datasets; simulating human behavior; making the expressions contained in the text they send, tone of voice words, and Internet buzzwords seem very real; however, deepfake can also use real images, backgrounds, sound effects, and other information to achieve the purpose of deceiving detection systems.

## 5. Conclusion

With the rise of online social networks, there are more ways than ever for people to buy things online. Online reviews about buying things or getting services have become a big way for users to share their thoughts. Because of this, manufacturers and merchants who often write spam reviews to promote or discredit the quality of certain products or services to make money or keep their reputation have raised concerns. The best way to solve this problem is through deep learning. Researchers have evaluated this research based on how they extract features from review datasets and how they extract features from review datasets, using different methods and strategies to find solutions to the problem. The goal of this paper is to do a deep and thorough analysis of recent comparative research on detecting spam. The goal of this paper is to do a deep and thorough analysis of recent comparative research on detecting spam.

In the future researchers may take the challenges detailed in these authors' works and design more powerful and adaptable deep learning models to create systems that can completely fight the first sort of misinformation, therefore reducing or even reversing the trend of the number of bogus comments.

## References

[1] Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. Journal of Big Data, 2(1), 1-24.

[2] Rao, S., Verma, A. K., & Bhatia, T. (2021). A review on social spam detection: challenges, open issues, and future directions. Expert Systems with Applications, 186, 115742.

[3] 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.

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

[5] Alom, Z., Carminati, B., & Ferrari, E. (2020). A deep learning model for Twitter spam detection. Online Social Networks and Media, 18, 100079.

[6] You, L., Peng, Q., Xiong, Z., He, D., Qiu, M., & Zhang, X. (2020). Integrating aspect analysis and local outlier factor for intelligent review spam detection. Future Generation Computer Systems, 102, 163-172.

[7] Jain, G., Sharma, M., & Agarwal, B. (2019). Optimizing semantic LSTM for spam detection. International Journal of Information Technology, 11(2), 239-250.

[8] Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In 2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON) (pp. 900-903). IEEE.

[9] Xue, H., Li, F., Seo, H., & Pluretti, R. (2015, August). Trust-aware review spam detection. In 2015 IEEE Trustcom/BigDataSE/ISPA (Vol. 1, pp. 726-733). IEEE.

[10] 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), 1053-1073.