

A Survey on Fake Review Detection Using Machine Learning

Sharad Dahate¹, Prof. Swati Soni²

^{1,2}Dept of CSE

²Professor, Dept of CSE

^{1,2}Takshila Institute of Engineering & Technology, Jabalpur, Madhya Pradesh, India.

Abstract- Presently, review sites are frequently confronted with the spread of wrong information, this could be done by an individual spammer or group spammers who compose fake reviews to either advertise or demean certain products that are available. This paper focuses on the detection of these fake reviews using sentiment analysis. Various data preprocessing techniques are used to convert the reviews to the proper format for analysis and for detection. The paper analyzed the methodology which is being used for fake review detection using machine learning and Deep Learning.

Keywords- Fake Review, Social Media, Machine Learning, LSTM, Bi-LSTM Deep Learning.

I. INTRODUCTION

Reviews are progressively used by individuals and groups for making decision with respect to purchase and for scalable marketing and design of the product. Appreciative/Complimentary opinions often mean better income, sale and fame for the specific business or individual selling that product and hence this gives a reason for individuals and/or group spammers to post phony reviews for the purpose of promotion or demotion of specific products. Such individuals that do such malpractice are called opinion spammers and their activities are called opinion spamming [1]. In the past few years, several techniques have been used to detect such opinion spammers. Most approaches that have been proposed rely on supervised machine learning techniques and on distinct characteristics.

A. Machine Learning Techniques

a. Naive Bayes: The NB classifiers are a family of “probabilistic classifier” supported by the application of the Bayes theorem. The NB calculates a group of chances by combining of values in a given dataset. Decision-making is quick [2, 3].

b. SVM: SVM is a supervised machine learning model which makes use of certain classification algorithms for two group

based classification issues. It is employed for regression and classification analysis. [4].

c. Maximum Entropy: Maximum entropy is a technique used to estimate the probability distributions from data given. When nothing is known the over-riding principle is that, the distribution must be as uniform as possible, that is, it must have maximal entropy [5].

d. K-Nearest Neighbor K-NN is known as a lazy learning algorithm which categorizes objects based on closest training since it is a nonparametric approach. Performance of this algorithm relies on various key factors, which included - distance measure, a similarity measure or k-value parameter. In the most basic way; it will be a positive or negative class. Generally a single number “k” is used that decides how many neighbors influence the classification [6].

e. Decision Tree Classification is done by splitting the criteria. The classification is done by sorting the attribute values in a tree like structure. The representation of an attribute in an instance to be classified is done by a node and the representation of outcomes of test is done by edges. [7].

1.1 Fake Review Characteristics: Following are some of the characteristics of fake review which can be used for fake review detection.

1.2

1. Very Few Details about Reviewer.
2. Relatively short and poorly written.
3. One sided reviews.
4. Full of product details.
5. Few helpful votes by other consumers.
6. Unverified purchase.
7. No Review history.
8. Very high or very low review score.
9. Extremely positive or negative reviews.

So by using above features one can be able to predict fake reviews in online platforms.

Below are some of the characteristics of Fake Review and real Review mentioned by [8]. The authors in [8] have clearly described some of the features of fake real reviews.

a. Many fake reviews are trying to establish credibility using “I” and “me” statements and overuse of verbs, therefore, creating the false impression that the reviewer has personal experience of the product or service.

b. Phrase repetition. If reviews use “like” or similar phrases over several reviews it can indicate falsehood as the fake reviewer may be producing bulk reviews for clients.

c. Due to the desire to seem credible, fake reviews set the scene using phrases like moving to a better lifestyle, a wealth creation business decision, my partner and I, again to create the impression of personal experience but with a nebulous surreal quality

d. One indicator of false reviews is the timing of reviews, if there is a spike in the number of reviews within a certain period of time or all reviews are very recent or dated prior to the release of a product then there should be some doubt about the veracity of the reviews

e. Fake reviews use generic names or faceless profiles sometimes because the reviews are from overseas marketing organisations

f. Incorrect spelling and grammar as many false reviews are written using poor English as they are generated by overseas marketing firms using people who are unfamiliar with our language

g. All black or all white, if the products has nothing but glowing reports with no drawbacks then the review may be questionable at best

h. Shorter reviews without considered discussions that are difficult to read may indicate falsehoods also

i. Reviews that have more subjective content or are based more on anecdotal evidence than facts, research or data are questionable whereas some detail about products or services is more likely characteristic of genuine reviews

j. Check out the date of the review especially if it is prior to the release of a product. This is a dead giveaway for falsehood.

k. Beware of reviews that hand out 5-star ratings for many products, if possible look into the reviewers profile to see if this happens consistently

l. Over the top or extreme enthusiasm about products is questionable as to its authenticity, Like the old saying goes ‘if it seems too good to be true it probably is’ Most genuine reviews point out a few drawbacks with a product or service

m. Competitor bashing or a competitor’s product denigrating at the same time as referring positively to another similar product with possibly a link to that product is highly spurious.

1.2 Characteristics of Truthful Reviews

a. Truth tellers use more nouns throughout their review like inventory, packing, parking, reliability, equipment and more about the facts than the feelings.

b. The scene-setting is irrelevant to giving a product a deserved recommendation.

c. Use of authenticity in people’s names and profiles and in particular a photo and verification of purchase or service use indicates genuineness.

d. Genuineness in phraseology and colloquialism in lingo. Use of real person language rather than industry babble which is a dead giveaway. Product users aren’t usually au fait with industry terms.

e. Greater spread of dates with some older and spikes of dates only when special sales occur indicates a more believable review.

f. Correct spelling and accurate use of grammar.

g. 3 stars or middle of the road reviews are more likely to be an honest evaluation of a product as most honest reviewers can see or experience the drawbacks of products as well as their benefits.

h. Grounded and balanced product endorsement rather than over-the-topedness. Unpaid testimonials usually are more balanced and realistic.

i. A logical flow of thoughts and just good common sense rather than pie-in-the-sky statements

j. Reviews that are based on a genuine and factual experience with the product or service with some data or research included

k. Believable and random dates of reviews

l. Balanced and fair evaluation of products without negative references to competition. Usually if there is any competitor bashing it is generated by a rival product or sponsor

m. If in doubt ask the reviewer themselves if contact details are available as usually fake reviewers will not respond to contact from enquirers

n. Check out whether the review is based on a verified or unverified purchase. This can establish the authenticity of the review. There could be a statement of the review being generated after a genuine purchase

II. LITERATURE REVIEW

Huayi Li et al. [1] reports a study of detecting fake reviews from Dianping (which has a built in fake review 8 system) Dianping classifies the reviews as positive (fake) and unlabeled (unknown) and hence they find the fake reviews from the unlabeled set. The PU-Learning learns from a set of positive and unlabeled Examples. They use The Spy Algorithm to identify some reliable negatives from the unlabelled set (U) EM using the naive bayes as the learning algorithm. It runs the NB (classifier) iteratively until it converges. Best accuracy: Spy + EM = 89.0%. Precision = TP/(FP+TP).

Julien fontanarava et al. [9] provided an in-depth analysis on the significant review and reviewer centric features that have been used to detect false reviews. The review centric features consist of: Density, Mean Rating deviation, early time frame, deviation from local mean.

Reviewer centric features: maximum content similarity, average content similarity and word number average and Algorithm, we augment some clipping statements to reduce the LCS algorithm's called times [10].

Somayeh Shojaee et al. [11] applied Stylometric features, using supervised machine learning classifiers, i.e. Support Vector Machine (SVM) with Sequential Minimal Optimization (SMO) and Naive Bayes, to detect deceptive opinion. Reviews are taken from 20 of the most commonly used hotels in Chicago from Trip Advisor and dishonest reviews collected using Amazon Mechanical Turk.

Qingxi Peng et al. [12] incorporated the sentiment analysis techniques into review spam detection and put forward a way to calculate sentiment score from the natural language text. They also compared different sentiment lexicon and found MPQA+Product got best accuracy of 61.4%.The comparison of different methods with different dataset the

sentiment score method had the best accuracy. Pan Liu et al. [13] presented a number of opinion spam detection's ID indicators on the basis of behavioral features of the individual. Methods they have used are conversion matching algorithm (CMA)- Its function is to show how many times a review can be converted to another review large common substring (LCS)-Works if the sentence structure of the review is changed by the spammer. Calculate Similarity Based on the LC.

IoannisDematis et al. [14] proposed an approach which integrates content and usage information to detect fake product reviews. Basic Spam Indicators- Rating Deviation (RD), Number of Reviews (NR), Content Similarity (CS). ChengaiSun et al. [15] have detected fake reviews of shops they have compared the accuracy of the different baseline algorithm they have used: BI-GRAMS (SVM), TRIGRAMS (SVM), PWCC (CNN classifier) and bagging (CNN+BI-GRAMS SVM+TRI-GRAMS SVM).The authors in [15], stated that, the bagging achieved the best performance accuracy of .781.

The author in [16] compares CNNs, GRUs and LSTMs. On NLP tasks: sentiment/ relation classification, textual entailment, answer selection, question-relation matching in Freebase, Freebase path query answering and part-of- speech tagging. The authors concluded, that, for Text, Sen tic GRU, they achieved an accuracy of 86.32% and RC GRU an accuracy of 68.56%. For Rematch, AS CNN got accuracy of 65.01%, QRM CNN got accuracy of 71.50%.

Y. Wang et al.[17] used an automatic keyword extraction method based on a bi-directional long short memory RNN. They filtered the review if they were relevant or not using an LSTM RNN based filter. If the review was relevant it was sent to the Keyword extraction stage. In keyword extraction stage LSTM RNN based classifier trained to map the original sentences into labels. Labels which occurred more often were used as keywords. The authors achieved the filtering words accuracy of 98.9% and keyword extraction foe words as 91.5%.

Alex Graves, et al. [18] made use of five neural network architectures in their work: Bidirectional LSTM, with two hidden LSTM layers. Unidirectional LSTM, with one hidden LSTM layer (LSTM).Bidirectional RNN (BRNN).Unidirectional RNN (RNN). Multi-Layer Perceptrons (MLP).The authors in [18] concluded that LSTM is comparatively faster and has a higher accuracy as compared to the others. [19] Chih-Chien Wang, et al. [20] attempted to use Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN) framework to detect spammers. Methods

used Network: long short-term memory (LSTM).The authors found that, the LSTM is more accurate than SVM and other traditional methods.

III. CONCLUSION

From the literature survey conducted on the number of research papers for sentiment analysis on various kinds of formats, we found that the most commonly used Machine Learning techniques were Naïve Bayes, decision Tree, Random forest and SVM classifier. However amongst these machine learning techniques used by the different authors of the various research papers, the most accurate is the Naïve Bayes model, with the highest accuracy is of 90.423% in the model used by Satuluri Vanaja. Meena Belwal [4] and the most common Deep Learning Techniques are the different models of RNN as LSTM, Bi- LSTM and GRNN where amongst the many deep learning techniques, the most accurate is the LSTM model in most cases with the highest accuracy is of 98.9% in the model used by Y. Wang, J. Zhang [16] for filtering words. Hence we will compare the combinations with each other and this will be implemented as a part of our further study

REFERENCES

- [1] L. Huayi Li, Bing Liu, Jidong Shao, "Spotting Fake Reviews using positive-unlabelled learning" (SciELO 2014).
- [2] Elshrif Elmurngi, Abdelouahed Gherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques".
- [3] Rakibul Hassan, Md. Rabiul Islam, "Detection of Fake Online Reviews Using Semi-supervised and supervised learning" (IEEE2019).
- [4] Satuluri Vanaja, Meena Belwal, "Aspect – Level Sentiment Analysis on E-Commerce Data" (IEEE – 2018).
- [5] Kamal Nigam, John Lafferty, Andrew McCallum, "Using Maximum Entropy for Text Classification".
- [6] Seema Sharma, Jitendra Agarwal, Shikha Agarwal, Sanjeev Sharma, "Machine Learning Techniques for Data Mining: A Survey".
- [7] Chetna Pujari, Aiswarya and Nisha P.Shetty, "Comparison of Classification Techniques for Sentiment Analysis of Product Review Data" (Springer – 2018).
- [8] <https://www.austate.com.au/blog/how-to-spot-fake-reviews>.
- [9] Julien Fontanarava, Gabriella Pasi, Marco Viviani, "Feature Analysis for fake review detection through supervised classification"(IEEE- 2017).
- [10]Pan Liu, Zhenning (Jimmy) Xu, Jun Ai, Fei Wang, "Identifying Indicators of Fake Reviews Based on Spammer’s Behaviour Features".
- [11]Somayeh Shojaee, Masrah Azrifah Azmi Muradt , Azreen Bin Azman, Nurfadhline Mohd Sharefl and Samaneh Nadali, "Detecting Deceptive Reviews Using Lexical and Syntactic Features"
- [12]Qingxi Peng and Ming Zhong, "Detecting Spam Review through Sentiment Analysis" (2014 ACADEMY PUBLISHER).
- [13]Pan Liu, Zhenning (Jimmy) Xu, Jun Ai, Fei Wang, "Identifying Indicators of Fake Reviews Based on Spammer’s Behaviour Features".
- [14]Ioannis Dematis, Eirini Karapistoli, and Athena Vakali, "Fake Review Detection via Exploitation of Spam Indicators and Reviewer Behaviour Characteristics".
- [15]Chengai Sun, Qiaolin Du,and Gang Tian, "Exploiting Product Review Features for Fake Review Detection".
- [16]Yuming Lin, Tao Zhu, Hao Wu, Jingwei Zhang, Xiaoling Wang, Aoying Zhou, "Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review sequence".
- [17]Y. Wang, J. Zhang, "Keyword extraction from online product reviews based on bi-directional LSTM recurrent neural network".
- [18]Wenpeng Yiny, Katharina Kanny, Mo Yuz and Hinrich Sch" utzey, "Comparative Study of CNN and RNN for Natural Language Processing".
- [19]Chih-Chien Wang, Min-Yuh Day, Chien-Chang Chen, Jia-Wei Liou, "Detecting Spamming Reviews Using Long Short-term Memory Recurrent Neural Network Framework".
- [20]Minlie Huang, Yujie Cao, Chao Dong, "Modeling Rich Contexts for Sentiment Classification with LTSM" (ResearchGate – 2016).