

Fake Online Reviews Detection Using Semi-Supervised And Supervised Learning

¹Kona Venkata Sai Mounica, ²D. Lalitha Bhaskari

¹M.tech Student, Department of Information Technology,

²Professor, Department of Computer Science and Systems Engineering,
Andhra University College of Engineering (A), Andhra University, Visakhapatnam, AP, India.

Abstract

Today's business and commerce is greatly impacted by Online reviews. These Online reviews are helping the users in making decisions regarding the purchase of online products. Online reviews are a great source of reputation for the companies. These reviews also have a great impact on promotion and advertisement of products and services. Hence some individuals or groups are trying to manipulate these online reviews for their own interests. This paper gives some semi-supervised and supervised text mining models to detect the fake online reviews and compares the efficiency in both the techniques on a dataset containing the reviews of some hotels.

Index terms:

Fake Online reviews, Semi- supervised learning, Supervised learning, Expectation Maximization algorithm, Decision tree classifier, Naive bayes classifier, Support Vector Machine classifier.

1) INTRODUCTION

A lot of advancement and change in technology is seen day by day. Old technologies are constantly being replaced by the new ones. These new technologies are helping the people to do their work efficiently. An example for such an evolution in technology is online marketplace. We can shop online and we can make reservations to buy the products in the online marketplace. Almost everyone of us who want to purchase an item or any particular thing will check the reviews of that particular thing in order to take a decision whether to buy or not. Hence these online reviews are a great source of reputation for the companies. These reviews also have a great impact on promotion and advertisement of products and services. As a part of these fake online reviews are also becoming a major concern. People can also make fake reviews to promote their own products which is a great harm for the other online business stake holders and product users. even competitive companies can damage the reputation of the other companies by writing negative fake reviews.

Many approaches have been studied by researchers for the detection of fake online reviews. Some are based on the content of the review and some are based on the behavior of the user who is posting the reviews. Content based approach focuses on the content of the review i.e, the text of the review where as the User behavior based approach focuses on number of posts that the reviewer made, country, ip-address etc. Most of the proposed approaches are supervised classification models. Some researchers have also worked with semi-supervised models. They are being introduced for the lack of reliable labeling of the reviews.

In this paper, A few supervised and semi-supervised classification approaches for detecting fake online reviews are studied and implemented. In semi-supervised learning, Expectation-maximization algorithm is used. Decision tree, Naive bayes and Support Vector Machines (SVM) are used as classifiers. The content based approach i.e, the content of the review is mainly focused on. As features sentiment polarity, frequency count and the length of the review is used.

2) RELATED WORK:

There are many techniques and approaches which are proposed in the field of fake review detection. The following approaches have been able to detect the fake online reviews with high accuracy. These techniques are divided into two categories.

a) Content Based Approach: This approach focuses on the content of the review. This can be text or anything which is told in the review. There were attempts made to detect the fake reviews by analyzing the linguistic features of the review. A researcher used three techniques to perform classification. These three techniques are Identification of genre, Psycholinguistic deception detection and Categorization of text.

- Identification of Genre: The parts-of -speech distribution of the review are explored. The frequency count of POS tags are used by them as the features representing the review for classification.
- Psycholinguistic Deception Detection: The Psycholinguistic approach assigns psycholinguistic meanings to the important features of a review.
- Categorization of Text: They experimented n-gram which is popularly used as an important feature in fake review detection

Other linguistic features are also explored. Some researchers took lexicalized and unlexicalized syntactic features by constructing sentence parse trees for the detection of fake reviews. They have shown that deep syntactic features improve the accuracy of prediction experimentally. Metadata of reviews such as date, time, length of the review and rating are also used by some researchers as features.

b)Behaviour Feature Based Approach: This approach focuses on the reviewer. i.e, the characteristics of a person giving the review. People who post fake online reviews intentionally have different behaviour compared to the normal user. They have the following deceptive review and rating behaviour.

- Giving unfair rating often: Generally Professional spammers post more fake reviews than normal ones. For example if the average rating of a product is 9.0 out of 10. But if a reviewer gives 3.0 rating. Analyzing the other reviews of that particular reviewer, we can detect him as a spammer.
- Giving good rating to their own country's product: People post deceptive reviews inorder to promote the products of their own region. This kind of spamming is mostly observed in case of movie reviews. For example, if there is an Indian movie in the International movie website and the average rating is given as 9.0 out of 10 and most of the reviewers are from India. This kind of spamming can be identified using the address of the reviewers.
- Giving reviews to different types of products: Each person has a specific interest of his own. A person generally will not have interest in all kinds of products. For example, a person who is interested in gaming will not be interested in classic literature. But if we find some people who gives reviews to different types of products which exceeds the general behaviour then we can say that they are making intentional fake reviews.

Detection of fake online reviews is generally considered as a classification problem and the most popular approach is to use supervised text classification techniques. These techniques are robust if large datasets of labeled instances from both classes, truthful opinions and deceptive opinions are taken for training. Some researchers used semi-supervised classification techniques as well.

Supervised learning has some challenges to overcome. The following are the problems with supervised learning.

- It is difficult to achieve the assurance of the quality of reviews.
- It is difficult to obtain the labeled data points to train the classifier.
- Humans are not too good in labeling reviews as fake or genuine.

Semi supervised method has been proposed where both labeled and unlabeled data are trained together. They proposed semi-supervised method in the following situations.

1. When there is no availability of reliable data.
2. The dynamic nature of online reviews.
3. Designing heuristic rules are difficult.

Several semi-supervised learning techniques are proposed which includes Expectation Maximization, Co-training, Label Propagation and Spreading and Positive Unlabeled Learning. Several classifiers like Random Forest, k-Nearest neighbor and Logistic Regression are used.

3) PROPOSED WORK

A) Dataset Description:

In this paper, a dataset have been used which consists of 1600 reviews, out of which there are 800 reviews of positive sentiment polarity and 800 reviews of negative sentiment polarity. Out of the 800 reviews of positive sentiment polarity, 400 are truthful reviews and 400 are fake reviews. Similarly the other 800 reviews of negative sentiment polarity also has 400 truthful reviews and 400 fake reviews. For the purpose of evaluations, a tag of '1' denotes truthful reviews and a tag of '0' denotes fake reviews. The reviews are collected from kaggle. It has reviews from different sources like MTurk, TripAdvisor, Web etc.

B) Proposed Methodology:

For the fake online reviews detection, we start with the raw data. A dataset which is labeled by the previous researchers is used. Unnecessary text like prepositions and articles are removed. Then the text data was converted into numerical form to be suitable for the classifier. Most important features are extracted and the classification took place. Steps like removing inconsistency, removing redundancy, handling missing values etc are not required. Instead merging the texts, creating a dictionary and mapping of text to numerical values are needed as a part of preprocessing. Sentiment polarity, length of the review and frequency count are used as features. The process of feature extraction is shown in figure 1.

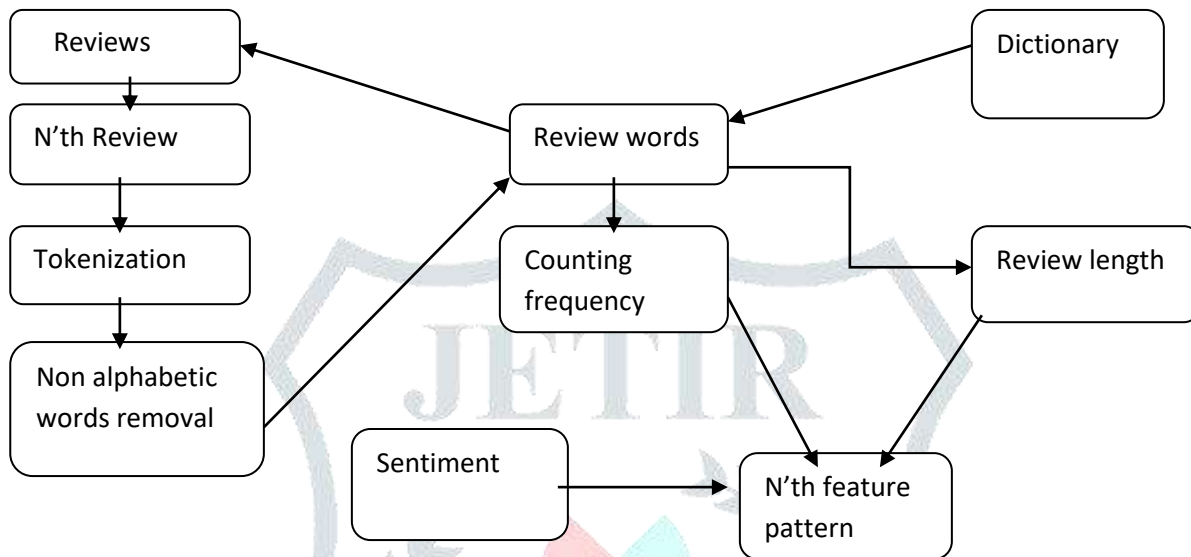


Figure 1

From the above figure, it is clear that we are working on the nth review, its features are generated in the following procedure.

1. Firstly, each review goes through tokenization process. Then, unnecessary words are removed and generation of candidate feature words takes place.
2. Each and every candidate feature word is checked against the dictionary and if the entry of it is available in the dictionary, its frequency is counted and is added to the column in the feature vector which corresponds to the numeric word of the map.
3. Along with counting frequency, The Reviews length is also measured and added to the feature vector.
4. Finally, the sentiment score present in the dataset is added in the feature vector. We have assigned zero to negative sentiment and positive valued in feature vector to positive sentiment.

Both supervised and semi-supervised classifications are implemented. For semi-supervised classification, we have used Expectation-Maximization (EM) algorithm. Expectation Maximization algorithm is designed to label unlabeled data to be used for training. Expectation Maximization algorithm is as given below.

Expectation Maximization algorithm**INPUT:** Labelled instance set L , and unlabelled instance set U .**OUTPUT:** Deployable classifier, C .

```

1:  $C \leftarrow \text{train}(L)$ ; //Derivation of classifier from the labeled dataset
2:  $PU = \emptyset$ ; //Initially null value is assigned to PU
3: while true do
4:  $PU = \text{predict}(C, U)$ ; // labeling of unlabelled data
5: if  $PU$  same as in previous iteration then
6:   return  $C$ ; //returning the classifier
7: end if
8:  $C \leftarrow \text{train}(L \cup PU)$ ; //Derivation of classifier from the combined set of labeled and unlabeled data
9: end while

```

The classifier is derived from the labeled dataset. That classifier is then used to label the unlabeled data. Let the predicted set of labels be PU . Then another classifier is derived from the combined sets of both labeled and unlabeled data and is used to classify the unlabeled data again. This process is repeated until PU gets stabilized. After getting a stable PU set, we trained the classification algorithm with the combined set of both labeled and unlabeled data and deploy it for predicting test dataset. In simpler words algorithm works in following two steps:

E-step

Estimates the expectation of the missing value i.e, the unlabeled class information. This step corresponds to performing classification of each unlabeled document. Probability distribution is calculated using current parameter.

M-step

Maximizes the likelihood of the model parameter using the previous computed expectation of the missing values as if were the true ones.

Support Vector Machine (SVM), Naive Bayes(NB), Decision tree classifiers with Expectation Maximization algorithm are used in this paper. The sophisticated library of these classifiers is provided by Scikit Learn package of Python programming language. Support Vector Machine, Naive bayes and Decision tree classifiers are used in Supervised learning. The highest accuracy is obtained for supervised Decision tree classifier. In addition to this we have developed a module in which the user can dynamically enter the review and prediction will be made on it.

A web application is created in which there will be Admin login, user registration and user login. All the accuracies of the different classifiers are shown in the Admin login in their respective techniques i.e, the supervised or semi-supervised learning. In the User login, the user will be able to give the review dynamically which will be replicated in the admin's login for the prediction and when clicked on predict, the review will be predicted using Supervised Decision tree classifier as that got the highest accuracy when tested with the dataset. The result will be sent to the users login and any user who logs in will be able to see all the reviews entered by the users and their results.

Detection of fake online reviews using supervised and semi supervised learning		
MyDetails Product Your Review View Reviews Logout		
user entered reviews		
ID/Item	Review	Predicted Value
1000000001	This hotel is so good!	1 Review is a useful review
1000000002	This hotel is the perfect location for shopping. Change shopping. The only thing is the food is extremely small - it is instant, but looks much larger on the website.	0 Predicted

Fig. 2 Prediction of User entered reviews

4) RESULTS AND PERFORMANCE ANALYSIS

A) Results

Expectation Maximization algorithm is implemented for semi-supervised classification in this paper. As classifiers Support Vector Machine, Naive bayes and Decision tree classifier are used. For semi supervised classification with SVM , an accuracy of 62.66% is found , with Naïve bayes an accuracy of 62.79% is found and with Decision tree classifier an accuracy of 80.66% is found.

Supervised classification techniques are also tried out to know the performance of our dataset. For supervised classification with SVM, an accuracy of 66.33% is found, with Naïve bayes an accuracy of 78.29% and with Decision tree classifier got an accuracy as 90.66%.

Apart from this, another module is developed in which users can dynamically enter the reviews and they will be predicted as fake or truthful based on supervised Decision tree classifier as that got the highest accuracy.

B) Performance Analysis

The performances of our implemented techniques are shown using the histogram below.

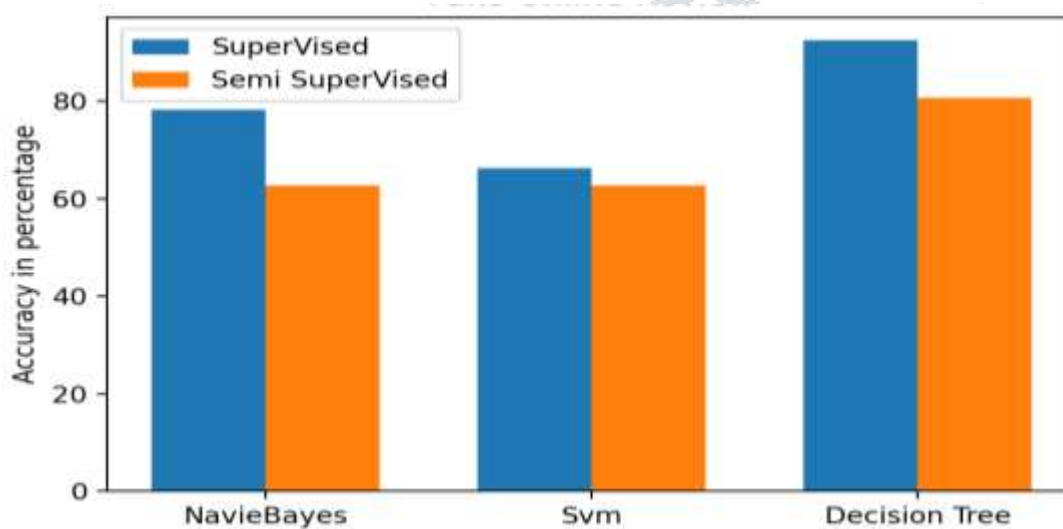


Fig.3 Histogram showing the performances of the implemented techniques

In this research, features are carefully chosen in order to reduce over fitting. We took the length of the review as a feature as it has a great significance. Highest accuracy of 90.66% is found by using Supervised classification with Decision tree classifier. The findings are summarized in the following table1.

COMPARATIVE SUMMARY OF SEMI-SUPERVISED AND SUPERVISED LEARNING TECHNIQUES ON HOTEL REVIEWS DATASET

Features	Algorithm type	Classifier used	Accuracy
Word frequency count, Sentiment Score, Length of the review	Semi-supervised	Decision Tree	80.66
		Naive Bayes	62.79
		SVM	62.66
	Supervised	Decision Tree	90.66
		Naive Bayes	78.29
		SVM	66.33

Table 1

5) CONCLUSION AND FUTURE WORK:

In this work three semi-supervised and supervised classification techniques for the detection of fake online reviews are implemented. In order to create a better feature set, features from several research works are combined. The highest accuracy is found in Supervised Decision tree classifier. In this paper, we have just worked on the reviews. User behaviour can also be combined with texts to have a better model for classification. The evaluation of the effectiveness of the proposed methodology can be done for a larger data set.

REFERENCES:

- [1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, 2016.
- [2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, "Detection of review spam: a survey", Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.
- [3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.
- [4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic Inquiry and Word Count: Liwc," vol. 71, 2001.
- [5] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.
- [6] J. Li, M. Ott, C. Cardie, and E. Hovy, "Towards a general rule for identifying deceptive opinion spam," in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
- [7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
- [8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, "Revisiting semi-supervised learning for online deceptive review detection," IEEE Access, Vol. 5, pp. 1319–1327, 2017.
- [9] <https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus>

