

Fake Detection of Online Reviews using Semi-Supervised and Supervised Learning

Aneel Narayanapur^{*1}, Pavankumar Naik², Suraksha G³, Pavitra S I⁴, Shruddha Mudigoudar⁵, Megha Honnali⁶

^{1,2,3,4,5,6}Department of Computer Science and Engineering, SKSVMACET Laxmeshwar – Karnataka, India

ABSTRACT

Online reviews have great impact on today's business and commerce. Decision making for purchase of online products mostly depends on reviews given by the users. Hence, opportunistic individuals or groups try to manipulate product reviews for their own interests. This paper introduces some semi-supervised and supervised text mining models to detect fake online reviews as well as compares the efficiency of both techniques on data set containing hotel reviews.

Keywords : Fake reviews, semi-supervised learning, supervised learning, Naive Bayes classifier, Support Vector Machine classifier, Expectation-maximization algorithm.

I. INTRODUCTION

The Technologies are changing rapidly. Old technologies are continuously being replaced by new and sophisticated ones. These new technologies are enabling people to have their work done efficiently. Such an evolution of technology is online marketplace. We can shop and make reservation using online websites. Almost, everyone of us checks out reviews before purchasing some products or services. Hence, online reviews have become a great source of reputation for the companies. Also, they have large impact on advertisement and promotion of products and services. With the spread of online marketplace, fake online reviews are becoming great matter of concern. People can make false reviews for promotion of their own products that harms the actual users. Also, competitive companies can try to damage each others reputation by providing fake negative reviews.

Researchers have been studying about many approaches for detection of these fake online reviews.

Some approaches are review content based and some are based on behavior of the user who is posting reviews. Content based study focuses on what is written on the review that is the text of the review where user behavior based method focuses on country, ip address, number of posts of the reviewer etc. Most of the proposed approaches are supervised classification models. Few researchers, also have worked with semi-supervised models. Semi-supervised methods are being introduced for lack of reliable labeling of the reviews.

In this paper, we make some classification approaches for detecting fake online reviews, some of which are semi-supervised and others are supervised. For semi-supervised learning, we use Expectation-maximization algorithm. Statistical Naive Bayes classifier and Support Vector Machines(SVM) are used as classifiers in our research work to improve the performance of classification. We have mainly focused on the content of the review based approaches. As feature we have used word frequency count, sentiment polarity and length of review.

II. METHODS AND MATERIAL

For detection of fake online reviews, we start with raw text data. We have used a data set which was already labeled by the previous researchers. We remove unnecessary texts like article and prepositions in the data. Then these text data are converted into numeric data for making them suitable for the classifier. Important and necessary features are extracted and then classification process took place.

As we have used 'gold standard' data set prepared by Ott et al. [3], we did not require the steps like handling missing values, removing inconsistency, removing redundancy etc. In stead we needed to merge the texts, create a dictionary and map the texts to numeric value as the tasks of pre processing.

we have used word frequency count, sentiment polarity and length of the review as our features. We have taken 2000 words as features. Hence the size of our feature vector is 160 2002. We have not taken n-gram or parts of speech as features because these are the derived features from bag of words and may cause over-fitting. The process of feature extraction is summarized in the figure 1.

From the figure 1, we can see that, when we are working with i'th review, it's corresponding features are generated in the following procedure.

Each review goes through tokenization process first. Then, unnecessary words are removed and candidate feature words are generated.

Each candidate feature words are checked against the dictionary and if it's entry is available in the dictionary then it's frequency is counted and added to the column in the feature vector that corresponds the numeric map of the word.

Alongside with counting frequency, The length of the review is measured and added to the feature vector.

Finally, sentiment score which is available in the data set is added in the feature vector. We have assigned negative sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

We have implemented both semi-supervised and supervised classifications. For semi-supervised classification of the data set, we have used Expectation-Maximization(EM) algorithm. The Expectation Maximization algorithm, first proposed by Karimpour et al. [9], is designed to label unlabeled data to be used for training. The algorithm operates as follows: A classifier is first derived from the labeled dataset. This classifier is then used to label the unlabeled dataset. Let this predicted set of labels be PU. Now, another classifier is derived from the combined sets of both labeled and unlabeled datasets and is used to classify the unlabeled dataset again. This process is repeated until the set PU stabilizes. After a stable PU set is produced, we have trained the classification algorithm with the combined training set of both labeled and unlabeled datasets and deploy it for predicting test dataset [8]. The algorithm is given below.

As classifier, we have used Support Vector machines(SVM) and Naive Bayes(NB) classifier with EM algorithm. Scikit Learn package of Python programming language provides sophisticated library of these classifiers. Hence for our research work, we have used Python with scikit-learn and numpy packages. We have tuned the parameters of the SVM for better results. For supervised classification, we have used Naive Bayes and SVM classifiers. We know, Naive Bayes classifier can be implemented where conditional independence property is maintained. As,

text comes randomly from user mind, we can't know what the next line and word is going to be. Hence, Naive Bayes classifier is popularly used in text mining. It is probabilistic method hence it can be used both for classification and regression. It is also very fast to calculate.

III. RESULTS AND DISCUSSION

We have used Expectation maximization(EM) algorithm for semi-supervised classification. As classifier we have used Support Vector machines(SVM) and Naive Bayes classifier. We have divided our dataset into a train test ratio of 75:25 and 80:20 for each classification process.

For semi-supervised classification with SVM, we have tuned different gamma parameters keeping C parameter constant. The percentage accuracy graph is shown in the figure 3. From the graph we can see, for semi-supervised classification with SVM classifier, We have found an accuracy of 81.34% for 80:20 split ratio and 80.47% for 75:25 split ratio with gamma equal 0.3 and 0.6 respectively. For semi-supervised classification with Naive Bayes classifier we have got an accuracy of 85.21% and 84.87% respectively for split ratio of 80:20 and 75:25.

Jiten et al. [8] using semi-supervised classification with EM and Positively Unlabeled learning respectively, got high-est accuracy of 83.00% and 83.75% for train test ratio of 80:20. They have tried Logistic regression, K-nearest neighbor, Stochastic Gradient Descent and Random Forest as classifier.

We have also tried supervised classification techniques to find out performance of them for our dataset. We have used Naive Bayes and SVM classifiers. For SVM classifier we have tuned gamma parameter keeping C parameter constant for having a better fit of the model. The results are shown in the following figure 4.

For supervised classification with SVM classifier, We have found an accuracy of 82.28% for 80:20 split ratio and 82.04% for 75:25 split ratio with gamma equal 0.1 and 0.8 respectively. For supervised classification with Naive Bayes classifier we have got the highest accuracy of 86.32% and 86.21% respectively for split ratio of 80:20 and 75:25.

IV. CONCLUSION

Hence there are numerous character check techniques accessible today nobody can recognize all cheats altogether while they are really occurring, they generally detect it till the fraud has been found. This happens on the grounds that an exceptionally little number of exchanges from the all-out exchanges are deceitful in nature. SMOTE technique will find the fraud detection by sorting both normal transaction and fraud transaction. Credit card fraud We have shown several semi-supervised and supervised text mining techniques for detecting fake online reviews in this research. We have combined features from several research works to create a better feature set. Also we have tried some other classifier that were not used on the previous work. Thus, we have been able to increase the accuracy of previous semi-supervised techniques done by Jiten et al. [8]. We have also found out that supervised Naive Bayes classifier gives the highest accuracy. This ensures that our dataset is labeled well, as we know semi-supervised model works well when reliable labeling is not available.

In our research work we have worked on just user reviews. In future, user behaviors can be combined with texts to construct a better model for classification. Advanced preprocessing tools for tokenization can be used to make the dataset more precise. Evaluation of the effectiveness of the proposed methodology can be done for a larger data set. This research work is being done only for English

reviews. It can be done for Bangla and several other languages.

deceptive review detection,” *IEEE Access*, Vol. 5, pp. 1319–1327, 2017.

V. 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 echnologies .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

Cite this article as :

Aneel Narayanapur, Pavankumar Naik, Suraksha G, Pavitra S I, Shruddha Mudigoudar, Megha Honnali, "Fake Detection of Online Reviews using Semi-Supervised and Supervised Learning ", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 6 Issue 3, pp. 428-431, May-June 2020.
Journal URL : <http://ijsrcseit.com/CSEIT2063112>