Fake online review detection with semi-supervised and managed learning

KAKKERA PRASANTHI ASST.PROFESSOR, Narasaraopeta Institute of Technology

AREMANDLA SAI PUJITHA ASST.PROFESSOR Narasaraopeta Institute of Technology

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 dataset containing hotel reviews.

Key words: Online Reviews, Machine, Learning, classification, semi supervised, supervised.

INTRODUCTION

Technologies are quickly evolving. New and sophisticated old inventions are constantly substituted. This emerging innovations enable people to function effectively. An online marketplace is such a development of technology. Online web pages enable us to shop and make reservations. Nearly all of us tests certain goods or services before buying them. So, online feedback have been a reputable platform for businesses. They also have a major effect on goods and service advertising and promotion. Fake web reviews are increasingly of interest with the spread of the online marketplace. For marketing of their own goods, people may create fake reviews that hurt the users. Competitive businesses may even attempt, by false bad feedback to destroy each other's image.

Researchers have investigated several ways to spot these bogus web reviews. Certain methods are contents-driven review and others are based on the user's behavior. Content focused research relies on the text of the analysis that concentrates on the region, Ip-address, amount of articles of the reviewer, etc. The consumer behavioural approach focuses on country. Any of the methods presented are classification structures supervised. Few researchers have also focused on semi-controlled models. For lack of accurate marking of the reviews, half-supervised approaches are adopted.

In this article, we allow several classification methods, some of which are half-controlled and others are monitored, to identify fake online feedback. We use the Expectation Maximization Algorithm for semi-supervised learning. In our research work, statistical Naive Bayes classification and vector support machine(SVM) are utilized as classifiers to enhance classification efficiency. We concentrated primarily on the substance of the reviews. We used word rate count, feeling polarity and duration of examination as function.

LITERATURE REVIEW

The issue of false screening has been addressed since 2007[17]. In the Fake Feedback identification research two key types of features were exploited; textual and compartmental features. Textual characteristics apply to the oral essence of test operation. That is, textual characteristics mostly rely on the quality of reviews. Compartmental features apply to the nonverbal features of the examinations. They rely primarily on reviewer habits such as writing styles, emotional gestures, and the periods at which reviews are written. While it is complicated and critical to fix textual features, behavioral features are very relevant as well as have a high effect on the output of the bogus revision phase and cannot be overlooked. In some bogus analysis articles, textual features were widely seen. In [18], investigators employed controlled methods for the identification of false reviews. The SVM, Naive-bavies, KNN, K-star and Decision-Tree classificators are in usage. Three iterations of the dataset [8] of labelled film reviews [1400, 2000, and 10662 film reviews respectively performed simulation studies. In[9] the authors have detected the false feedback on the data collection they have obtained by using Naive Baye, Decision Treaties, SVM, the Random Forest and the Maximum Entropy Classificatory. The collected data collection contains approximately 10,000 derogatory tweets about Samsung's goods and services. In [20] both SVM and Naive basic categories were used by the writers. The authors have been working on the returns dataset, consisting of 1600 reviews from 20 famous Chicago hotels. In [21] the authors used the neural and discreet models of deep learning classificatory average, CNN, RNN, GRNN, mean GRNN and two-way average GRNN for detecting disappointing opinion spamming. The data collection from [12] included true and disappointing ratings in three areas: hotels, restaurants and physicians. Both these investigations have taken the textual characteristics only into account without attempt to comport.

In the false review identification mechanism, other publications called behavioral functions. In certain compartmental characteristics such as an overall ranking and the amount of reviews the reader wrote is taken into account in Amazon reviews. In an additional work[14], the authors studied the influence on the bogus assessment method of restaurant and hotel domain of both textual and behavioral characteristics. In[15] the proposal also includes the integration of textual and behavioral functionality through an iterative computing system plus plus (ICF++). They detected fake assessments focused on assessing the authenticity, trustworthiness and reliability of a commodity the evaluator

METHODOLOGY

The diagram suggested for the model as shown in figure 1 below. The sequence and flow of the experimental process followed by this work is presented here.

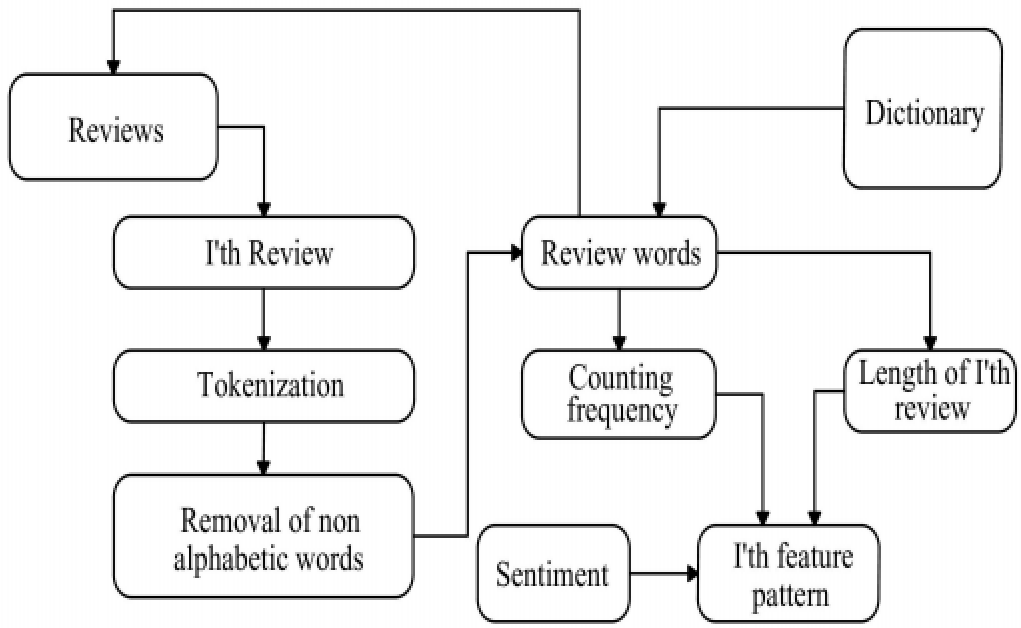

Fig.1: Feature Extraction process

We start with raw text data to identify fake online reviews. We used a dataset that the previous researchers had already named. We delete redundant texts such as articles and data prepositions. This text data is then translated to numerical data for classification. Important and essential features are extracted, followed by a method of classification. We did not need moves such as coping with missed values, extracting discrepecies, deleting duplication, etc., since we used a gold standard dataset prepared by Ott et al. » [3]. Instead we had the tasks of fusing the documents, creating a dictionary, and mapping the texts to numerical meaning. We used word count duration, feel polarity, and analysis duration as our characteristics. We took 2000 terms as characteristics. Therefore, our vector scale is 160 to 2002. We didn't use n-gram or half of the speech as characteristics, since these are the resulting characteristics of words pack and will over fit. Fig.1 summarizes the method of extracting functions.

From Figure 1, we can see that the corresponding functions are created in the following procedure when we work with the analysis

1) First of all, each review is tokenized. Subsequently redundant terms and nominee function words would be deleted.

2) Each candidate feature terms are tested against a dictionary and the number of the entry is counted and applied to the column in the feature vector which correlates to the number map of the term whether that entry is accessible in the dictionary.

3) The duration of the analysis is calculated and applied to the characteristic vector along with the counting frequency.

4) In the function variable is eventually inserted the sentiment score that is present in the data collection. In the function vector, we have provided a negative feeling as nil and a positive feeling.

Semi-supervised and supervised classifications have been introduced. We also used the Expectation-Maximization(EM) algorithm for the semi-supervised classification of the data collection. Karimpour et al. [9] suggested for the first time the algorithm for maximizing expectations is intended to mark unlabeled data for preparation. The algorithm works like: Second, a classifier comes from the data collection labelled. The unlabeled data collection will then be classified by this classifier. Allow that the forecast mark set to be PU. Now a further Classifier comes from both classified and unmarked data sets together and is used to reclassify the unmarked data set. This is done before the set PU stabilizes. Following production of a stable PU collection we have formed and deployed a hybrid qualified classification algorithm, both labelled and unlabeled, to predict the test dataset [8]. Below is the algorithm.

We used Support Vector(SVM) and Naive Bayes(NB) classificatory with EM algorithms to serve the classificatory. The advanced library of these classifiers offers Scikit Learn programming language package of Python. We have used Python with science-learning and various packages therefore for our study work. For better performance, we have specified the parameters of the SVM. We also used Naive Bayes and SVM classifiers for the supervised classification. We know that where the conditional independence property is preserved, Naive Bayes classification may be applied. Because text comes from the user's mind arbitrarily, we cannot understand the next line and phrase. Therefore, the classification Naive Bayes is commonly used in the mining of texts. It is probabilistic and can thus be used for classification as well as for regression. It can also be calculated very quickly**.**

RESULTS & EVALUATION

We used the semi-supervised classification Expectation Maximization(EM) algorithm. The Support Vector (SVM) and the Naive Bayes classification were used as a classificator. For each classification phase, we have divided our data set into a 75:25 and 80:20 ratios. We also tuned various gamma parameters to maintain the C parameter with semi-supervised classification with SVM. Figure 3 shows the percentage precision graph. The graph shows that we found a precision of 81,34 percent with an 80:20 split ration and 80,47 percent with a 75:25 divided ratio with a gamma equivalent to 0,3 and 0,6, respectively, for the semi-supervised classification with an SVM classification. We have a precision of 85.21% and 84.87% respectively for the semi-supervised classification with a controversial ratio of 80:20 and 75:25.

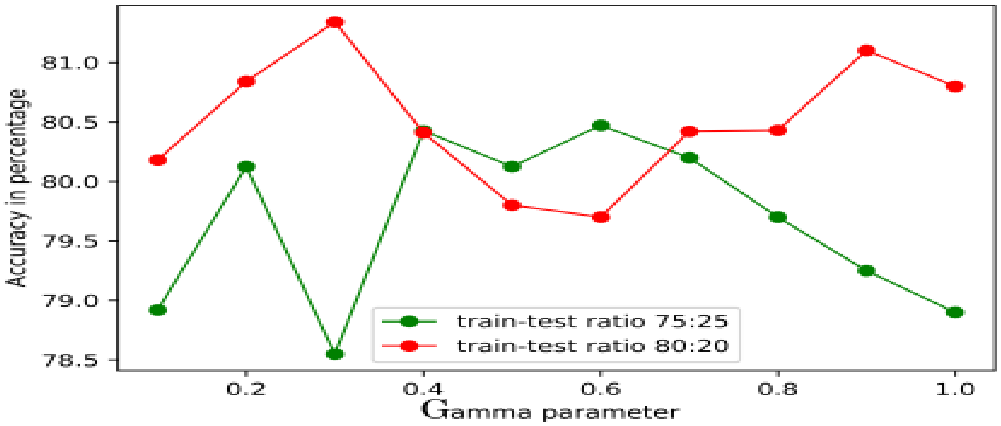

Fig. 2. Gamma parameter vs Accuracy for EM with SVM

We also tried to figure out the output of the supervised classification techniques for our results. We used SVM classifiers from Naive Bayes. We've specified the gamma parameter for the SVM classifier, holding the parameter C constant to match the model better. The corresponding figure 4 shows the findings.

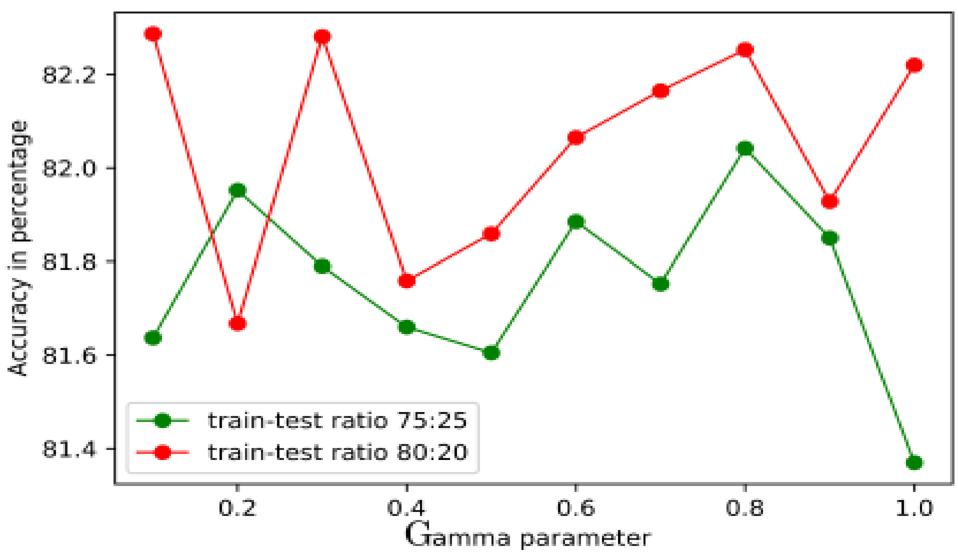

Fig. 3. Gamma parameter vs Accuracy for supervised SVM classier

We observed an accuracy of 82.28 percent with 80:20 split ratio and 82.04 percent with a 75:25 split ratio with a gamma equivalent to 0.1 and 0.8 respectively in supervised classification with SVM classification.

We have the best precision of 86.32 and 86.21 percent for the controlled classification for the Naive Bayes classifier for 80:20 and 75:25 divisions.

.

CONCLUSION

In this research we have seen many strategies of semi-supervised and monitored text mining to identify fake online feedback. For a greater collection of functionality, we merged features of many research works. We have also attempted another classification that has not been included in the previous work. We have thus been able to improve the precision of Jiten et alprevious.'s semi-supervised techniques[8]. We also noticed that the controlled classification Naive Bayes provides the utmost precision. This guarantees the etiquette of our data collection and that we know that the semi-controlled model is workable when accurate etiquette is not accessible.

We have just worked on consumer feedback of our field work. User habits may be mixed with texts to build a stronger classification model in future. Advanced tokenization preprocessing methods may be used to specify the data collection. For a broader dataset, evaluations may be made about the efficacy of the suggested technique. This study is conducted for English examinations only. For Bangladesh and many other languages it is likely.

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] J. Karimpour, A. A. Noroozi, and S. Alizadeh, "Web spam detection by learning from small labeled samples," International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.

[10] S. Shojaee et al., "Detecting deceptive reviews using lexical and syntactic features." 2013.

[11] Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: An empirical study," Information Sciences, vol. 385, pp. 213– 224, 2017.

[12] H. Li et al., "Spotting fake reviews via collective positive-unlabeled learning." 2014.

[13] N. Jindal and B. Liu, "Opinion spam and analysis," in Proceedings of the 2008 International Conference on Web Search and Data Mining, ser. WSDM '08, 2008, pp. 219–230.

[14] D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What online reviewer behaviors really matter? effects of verbal and nonverbal behaviors on detection of fake online reviews," Journal of Management Information Systems, vol. 33, no. 2, pp. 456–481, 2016.

[15] E. D. Wahyuni and A. Djunaidy, "Fake review detection from a product review using modified method of iterative computation framework." 2016.

[16] D. Michie, D. J. Spiegelhalter, C. Taylor et al., "Machine learning," Neural and Statistical Classification, vol. 13, 1994.

[17] T. O. Ayodele, "Types of machine learning algorithms," in New advances in machine learning. InTech, 2010.

[18] F. Sebastiani, "Machine learning in automated text categorization," ACM computing surveys (CSUR), vol. 34, no. 1, pp. 1–47, 2002.

[19] N. Jindal and B. Liu, "Review spam detection," in Proceedings of the 16th International Conference on World Wide Web, ser. WWW '07, 2007.

[20] E. Elmurngi and A. Gherbi, Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. IARIA/DATA ANALYTICS, 2017.

[21] V. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews and blog posts," in Advance Computing Conference (IACC), 2013, pp. 893–898.

[22] A. Molla, Y. Biadgie, and K.-A. Sohn, "Detecting Negative Deceptive Opinion from Tweets." in International Conference on Mobile and Wireless Technology. Singapore: Springer, 2017.

