
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 7, Issue 3, 2019

DETECTION OF FAKE ONLINE REVIEWS USING SEMI-SUPERVISED AND SUPERVISED LEARNING

Parvati Kadli, Manjula. S. D, Indira

Assoc. Professor, Assoc. Professor, Asst. Professor

kadli.parvati@gmail.com, manjulasd@pdit.ac.in, indira.raj.06@gmail.com

Department of CSE, Proudhadevaraya Institute of Technology, Abheraj Baldota Rd, Indiranagar, Hosapete, Karnataka-583225

ABSTRACT:

Reviews posted on the internet have a significant effect on modern trade and company. Customer reviews play a significant role in influencing consumers to buy things online. This leads some people or organisations to attempt to influence product evaluations in a way that benefits them. In this study, we present a few trained and semi-supervised text mining algorithms for detecting phoney reviews online and evaluate their performance on a dataset including hotel reviews.

Key words: Fake reviews, online reviews, Own interests.

I INTRODUCTION

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, every one 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 other's 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 behaviour 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 behaviour 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 labelling 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. In the following section II, we discuss about the related works. Section III describes our proposed approaches and experiment setup. Results and findings of our research are discussed in Section IV. Section V includes conclusions and future work.

2. RELATED STUDY

Social Web site and the increasing popularity of social media have resulted in the dissemination of many types of content (e.g. text, acoustic, visual) produced directly by users, so-called user-generated content (UGC). With Web 2.0 technology, it is possible for everyone to be able to use content on social media, almost without a reliable external control mechanism. This means that there are no means for verification, a priori, source credibility and credibility of the content produced. In this context, the issue of assessing the reliability of the data used by social media platforms is gaining increasing attention from researchers. In particular, this issue has been extensively investigated on review sites, where the distribution of inaccuracies in the type of spam, and the negative effects it poses, is extremely harmful to businesses and users. In this context, the detection of spam views aims to identify fake reviews, fake comments, fake blogs, misleading public posts, deceptions and misleading messages [1], and to make them easily known. Acquisition techniques for detecting non-targeted reviews are particularly on specific review sites such as TripAdvisor1 or Yelp, 2 where user reviews have a strong impact on people visiting the Website for advice. Therefore, a product or service recommendation such as a restaurant or hotel based on false information can have serious consequences. Many of the methods proposed to date to gain a partial overview on these forums rely on machine learning techniques that focus on unique features, i.e., features, linked to reviews and / or to the reviewers who have produced them. It has been shown in the literature that their use can lead to effective detection of suspicious content and / or reviewers, and due to false designations [2]. Recent methods have suggested the use of additional features that monitor the social composition of the network underlying the imaging review site. These methods, which are usually based on unsupervised graph manipulation methods, often provide the worst performance with respect to supervised solutions. On the other hand, supervised methods also present other issues. First, the solutions available tend to consider a small set of features, or different categories of features separately; Second, it was tested on small data extracted from well-known review sites previously. Therefore, the proposed solutions are for the most part partial, or site-dependent. Considering the various factors that have been proposed and used for the different monitoring methods, the purpose of this article is to provide a feature that reflects the most relevant and general features- and reviews-of the cents that can be used in the review area get a fake review. Among these features, some are well known and taken from books, some are new and create another paper. To test the use of this set of features in distinguishing real and fake reviews, a secure monitor has been based on a known machine learning process. As for the books, it is publicly viewed with big data from the Yelp.com review site. This allows to provide the most important results with regard to the contribution of each derived feature and the groups of features. In particular, an important contribution of a particular group of factors in analyzing the reliability of so-called singleton reviews has emerged. The reliable results obtained indicate the efficiency and application of the feature analysis shown in this article.

3 METHODOLOGY

Today, Sentiment analysis plays an important role where various machine learning technique is used in determining the sentiment of very huge amounts of text or speech. Various application tasks include such as determining how someone is excited for an upcoming movie, correlates different views for a political party with people's positive attitude towards vote for that party, or by converting written hotel reviews into 5-star based on scaling across categories like 'quality of food', 'services', 'living room' and 'facilities' provided. As there is huge amount of information is shared on social media, forums, blogs, newspaper etc. it is easy to see why there is a need for sentiment analysis

ISSN2321-2152
www.ijmece .com
Vol 7, Issue 3, 2019

as there is much information to process manually which is not possible in today's time.

As briefly introduced in Section II, many and different are the features that have been considered so far in the review site context to identify fake reviews. In some cases, features belonging to different classes have been considered separately by distinct approaches. In other cases, the employed features constitute a subset of the entire set of features that could be taken into account; furthermore, new additional features can be proposed and analyzed to tackle open issues not yet considered, for example the detection of singleton fake reviews. For these reasons, in this section we provide a global overview of the various features that can be employed to detect fake reviews. Both significant features taken from the literature and new features proposed in this article are considered. Since the most effective approaches discussed in the literature are in general supervised and consider reviewand reviewer-centric features, these two classes will be presented in the following sections. The choices behind the selection of the features belonging to the above mentioned classes will be detailed along each section. When the features are taken from the literature, they will be directly referred to the original paper where they have been initially proposed. The absence of the reference will denote those features that have been widely used by almost every proposed technique. Finally, the presence of the label denoted by [new] will indicate a feature proposed for the first time in this article. A. Reviewcentric Features The first class of features that have been considered, is constituted by those related to a review. They can be extracted both from the text constituting the review, i.e., textual features, and from meta-data connected to a review, i.e., metadata features. In every review site, the time information regarding the publication of the review, and the rating (within some numerical interval) about the reviewed business are metadata, are always provided. In addition, in relation to metadata features, those connected to the cardinality of the reviews written by a given user must be carefully studied. In fact, a large part of reviews are singletons, i.e., there is only one review written by a given reviewer in a certain period of time (this means that in the user account there is only one review at the time of the analysis). For this kind of reviews, specific features must be designed. In fact, as it will be illustrated in the following, many of the features that have been proposed in the literature are based on some statistics over several reviews written by the same reviewer. In the case of singletons, these features loose their relevance in assessing credibility. Therefore, the definition of suitable features that are effective for detecting also singleton fake reviews becomes crucial. 1) Textual Features: as briefly illustrated in Section II, it is practically impossible to distinguish between fake and genuine reviews by only reading their content. The analysis provided by Mukherjee et al. in [19] has shown that the KL-divergence between the languages employed by spammers and non spammers in Yelp is very subtle. However, the good results obtained in [26] by using linguistic features on a domain specific dataset (i.e., a Yelp's dataset containing only New York japanese restaurants), show that at least on a domain specific level, textual features can be useful. It is possible to use Natural Language Processing techniques to extract simple features from the text, and to use as features some statistics and some sentiment estimations connected to the use of the words.

4 RESULTS EXPLANATION

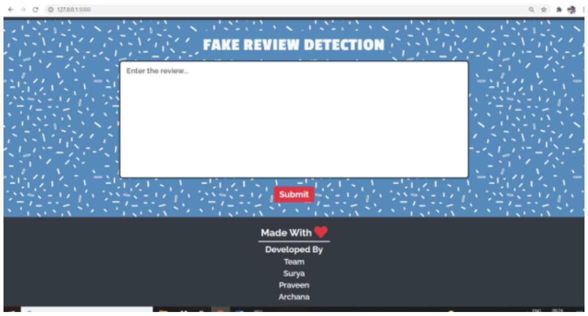

Fig.4.1. Home page.

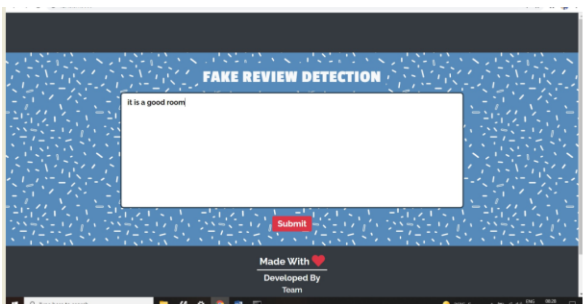

Fig.4.2. Review input.

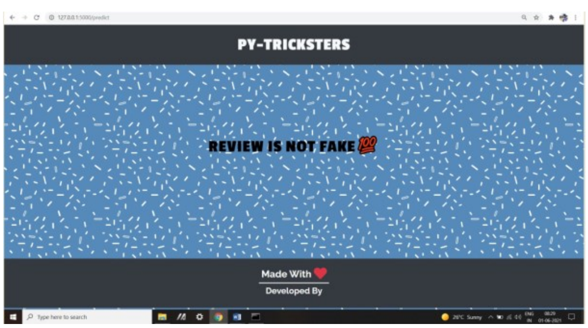

Fig.4.3. Not fake indication.

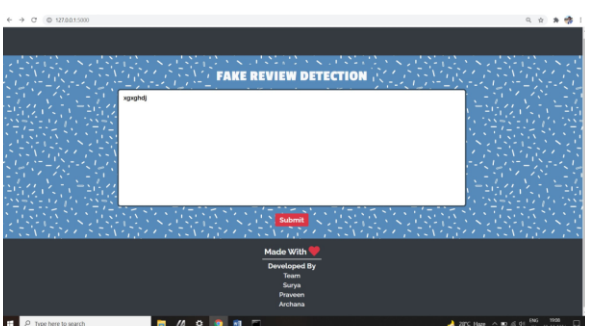

Fig.4.4. Input image.

Fig.4.5. Fake review detected.

CONCLUSION

In order to identify phoney reviews posted on the internet, we have shown a number of supervised and semi-supervised text mining algorithms. To improve the feature set, we incorporated features from many research studies. There were additional classifiers that were not part of the earlier study that we tested as well. As a result, we improved upon earlier semi-supervised methods developed by Jiten et al. [8]. Additionally, we discovered that the most accurate classifier is the supervised Naive Bayes classifier. This guarantees that our dataset has accurate labels, as semi-supervised models perform admirably in situations when trustworthy labelling is unavailable. We have focused only on user reviews in our study. In the future, it will be possible to build a more accurate categorisation model by combining user behaviours with texts. The dataset may be made more exact by using advanced pre-processing methods for tokenisation. With more data, we can see how well the suggested technique works. Only English reviews are being considered for this study. Bangla and a number of other languages can have this done.

REFERANCES

ISSN2321-2152
www.ijmece .com
Vol 7, Issue 3, 2019

[1] B. Wang, Y. Min, Y. Huang, X. Li, F. Wu, " Review rating prediction based on the content and weighting strong social relation of reviewers," in Proceedings of the 2013 international workshop of Mining unstructured big data using natural language processing, ACM. 2013, pp. 23-30.

[2] D. Tang, Q. Bing, T. Liu, "Learning semantic representations of users and products for document level sentiment classification," in Proc. 53th Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, July 26-31, 2015, pp. 1014–1023.

[3] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, S. Ma, "Explicit factor models for explainable recommendation based on phrase-level sentiment analysis," in proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 2014.

[4] W. Zhang, G. Ding, L. Chen, C. Li , and C. Zhang, " Generating virtual ratings from Chinese reviews to augment online recommendations," ACM TIST, vol.4, no.1. 2013, pp. 1-17.

[5] X. Lei, and X. Qian, "Rating prediction via exploring service reputation," 2015 IEEE 17th International Workshop on Multimedia

[6] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks, " in Proc. 18th ACM SIGKDD Int. Conf. KDD, New York, NY, USA, Aug. 2012, pp. 1267–1275.

[7] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," in proc. 21st ACM Int. CIKM, 2012, pp. 45-54.

[8] Z. Fu, X. Sun, Q. Liu, et al., "Achieving Efficient Cloud Search Services: Multi-Keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," IEICE Transactions on Communications, 2015, 98(1):190-200.

[9] Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual Verifiable Provable Data Auditing in Public Cloud Storage," Journal of Internet Technology, vol. 16, no. 2, 2015, pp. 317-323.

[10] W. Luo, F. Zhuang, X. Cheng, Q. H, Z. Shi, "Ratable aspects over sentiments: predicting ratings for unrated reviews," IEEE International Conference on Data Mining (ICDM), 2014, pp. 380-389