International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 06 | June -2020 ISSN: 2582-3930

DETECTION OF ONLINE REVIEWS USING SUPERVISED AND SEMI-SUPERVISED LEARNING

Mr.MD.ShakeelAhmed[1], P. LakshmiChandana[2], SK.Taranum[3], M. Sai Pranathi[4]

[1]Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

[2, 3, 4 ,5] Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, 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 online reviews as well as compares the efficiency of both techniques on dataset containing hotel reviews.

1. 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 reputation by providing negative reviews.

2.Related Work

The scope of the project is providing better hotels to the users based on user reviews subjected to different categories such as atmosphere, food, services etc.,. In this project we are using expectation maximization algorithm, which can fetch the user reviews and disperse it according to category based on keywords such as chicken, food, nice, delicious and so on. After dispersing the reviews, measure the positive

and negative reviews according to category then only can get the average rating of the particular hotels.

3.Existing System

In existing system many approaches are there for detection of these 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.

4. Proposed System

In this paper, we make some classification approaches for detecting 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.

5. Implementation

USER INTERFACE DESIGN

This is the first module of our project. In this the application user's first create their account properly which are stored at the back end for verification or for providing security to the accounts. If user wants to get into his account first they have to submit their constraints such as username, password and so on…otherwise can't able to access the account. In our project according to actions they are performing we disperse the users as admin or normal application user.

HOTEL:

I this project hotels will provide information about each individual like hotel name, city, Items available in that hotel. And as well as each hotel having their own reviews which are given by different users.

USER

In this project the user was searching for the hotels and he/she have a permission to upload a review on hotel as well as view the reviews on hotel. And also have permission for view review percentage of each hotel or all hotels at a time.

User was having following operations:

a. Register.
b. Login.
c. View hotels.
d. View reviews.
e. Give reviews.
f. Logout.

ADMIN:

In this project Admin can add the hotels into site as well as delete the hotels from site. And also admin allowing to view the hotels list and also reviews of hotels.

Admin have following operations:

a. Login.
b. View hotels.
c. View reviews.
d. Add hotels.
e. Delete Hotels.
f. Logout.

EMA:

EMA (Expectation Maximization Algorithm) is a framework used for aspect category detections from hotels reviews and differentiate the good and bad reviews from each hotel. And also give the percentage of good and bad reviews for each hotel.

We have implemented both semi-supervised and supervised classifications. For semi-supervised classification of the data set, we have used Expectation-Maximization (EM) algorithm.

ALGORITHM USED

Expectation-Maximization Algorithm

6. Conclusion and Future Enhancements

In this project we have shown several semi-supervised and supervised text mining techniques for detecting 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 Bangal and several other languages.

7. 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.

Clustering ", IEEE Computer Society Washington, DC, USA 2015.

[5] Luis M. de Campos, Juan M. Fernández-Luna *, Juan F. Huete, Miguel A. Rueda-Morales; "Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks", International Journal of Approximate Reasoning, revised 2010.

[6] UrszulaKużelewska; "Clustering Algorithms in Hybrid Recommender System on MovieLens Data", Studies in Logic, Grammar and Rhetoric, 2014.

[7] DietmarJannach, Gerhard Friedrich; "Tutorial: RecommenderSystems",InternationalJoint Conference on Artificial Intelligence, Beijing, August 4, 2013.

[8] Gaurangi, Eyrun, Nan; "MovieGEN: A Movie Recommendation System", UCSB.

[9] Harpreet Kaur Virk, Er. Maninder Singh," Analysis and Design of Hybrid Online Movie Recommender System"International Journal of Innovations in Engineering and Technology (IJIET)Volume 5 Issue 2,April 2015.