Detection of Online Fake Reviews Based on Text Emotion

S Kiruthika¹, V Vishnu Priyan²
¹Assistant professor, Department of Computer Science Engineering, Sri Krishna College of Technology, Coimbatore
²Student, Department of Computer Science Engineering, Sri Krishna College of Technology, Coimbatore
Kiruthika.s@skct.edu.in, 17tucs254@skct.edu.in

Abstract. Online groups offer clients with ways to overpower a few data hurdles and limitations, like the challenge to get self-governing data about movies and for the co-occurrence of positive and negative reactions within reviews. False reviews will disturb such choices because of misleading data, causing commercial disadvantages for the customers. Recognition of false opinions has thus expected very large concentration now days. But, many websites have only concentrated on handling the problematical comments and reviews. Our work tends to categorize people opinions into groups of true or false polarization by employing text feature analysis. In our work, our team analyze people opinions by implementing Sentiment Analysis techniques to recognize false opinions. Sentiment Analysis and text feature categorization techniques were used to a database containing people opinions. Furthermore, the estimation reviews acquired from reviewers could be categorized into good or bad opinions, that could be utilized by a customer to choose a movie. Further the proposed technique will graph based on the classification of true and fake reviews as the analysis of good and bad reviews for a product (movie). It will help us to predict the ratio of fake reviews to true reviews easily. To estimate the accomplishment of SA technique, this work has employed accurateness, exactness, recollection and F-degree as a performance rate.

Keywords: Sentiment Analysis, text feature analysis, positive and negative reviews, Graph analysis.

1. INTRODUCTION

The feeling Analysis has turn out to be one among the highly thought-provoking areas in the estimation of text features, thanks to its hopeful industrial profits. One among the chief challenges fronting reserves is the way to mine reactions at intervals the review, and the way to identify fake yet good reviews and fake yet bad opinions from estimation reviews. SA is one among the necessary areas. SA detects the character of a given review, because it is good or bad. Precisely, the objectives of this technique are to locate sentiments from opinions and then categorize this emotion related to polarization. There are three main categorizations. Thus, it is significant to differentiate various levels of an estimation method which will estimate the various functions. The first level reflects that a review is an view on its expression, and it targets to categorize a review as a bad or good view. The second level targets to frame review
indicated in each and every single sentence. Distinguishing such false reviews is a major problem. Our work is primarily focused to Sentiment Analysis at the document level, more precisely, on database of movie reviews. This technique and SA implementation are anticipated to provide a huge positive result, specifically for the identification of fakeness in updated movie reviews.

**Objective**

- To classify the updated movie reviews as negative and positive reviews.
- To detect the fake review in order to provide customers the true information about the product i.e., movie.

2. **LITERATURE SURVEY**

Deficiency of entrust in online firms is a prime cause for many users to hesitate to shop online. This work proposes a technique to describe to develop initial trust in new customers of a web-based company following their initial step. The technique is empirically verified by a questionnaire-based field analysis [1].

The kind of self-sufficiency and frankness of business in online social networks (ECOS) establishes an issue to the protection of contracts as it is problematic to guarantee the credibility of individuals on both sides. Entrust managing systems, logically, have arisen as possible resolutions [2].

There are various challenges to count while using SA. In this work, two chief challenges are focused. Initially, the perspective perceived as fake/false in a incidence might be reflected as true in another incidence. Then, publics usually avoid open-up their expressions in the similar manner. Many text processing techniques implement the idea that slight variations between the two text fractions are questionable to vary the original meaning [3].

Moraes et al. proposed a method for classifying a single subject word-based review. A sentiment classified is implemented for identifying a fake or true review. Supervised learning techniques consist of two stages, i.e., collection and abstraction of reviews utilizing learning models [4].

These reviews are accessible for clients to recite. Clients are progressively based on reviews. Character prototypes can utilize SA techniques to abstract clients’ expressions and utilize this data in the Character system. This data comprises clients’ expressions about various ideas [5].

3. **EXISTING METHOD**

Investigators have been analyzing about various techniques for identification of these false opinions. A few methods are opinion based and a few are related to performance of the reviewer [6]. Opinion related analysis concentrates on content of the opinion where reviewer performance related methodology concentrates on reviewres' personal details. Opinions became a huge supply of standing for the industries [7].

**DISADVANTAGES:**

- In this existing work, the reviews are classified related to the presence of problematic and offensive contents in the opinion, but this type of classification would fail in detecting the fake reviews which is not necessary to have offensive or problematic words.
4. PROPOSED TECHNIQUE

Some organization methods for estimating false opinions, among which partial proctored and others are fully proctored. They used frequency count of POS tags as the structures representing the review for classification [8]. In this technique, we used text feature extraction and analysis and also sentiment analysis technique to estimate the presence of good and bad opinions in movie review section. Also, the technique plot a graphical representation to enable us to understand the categorization more effectively [9].

ADVANTAGES:

- Since our proposed technique implements sentiment analysis for the classification and detection of true and fake reviews, it outperforms the other existing methods.
- Also, the graphical representation of the classified reviews will make the system highly user-friendly to predict the nature of the review.

5. SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

System: I5 processor
RAM: 4 GB
Monitor: 20" Color
Mouse: Logitech

SOFTWARE REQUIREMENTS:

OS: Windows 8
Language: python
Data set type: EXCEL
Front End: GUI

6. ARCHITECTURE DESIGN
Our system consists of two modes namely Admin mode and User mode. In Admin mode, the admin has to enter the username and password to login [10]. After logging in, the admin can update and view movie details. Also, the admin will view the reviews updated by the reviewers. In User mode, the user will enter the username and password to login if he/she is a registered user. Else, he/she has to register themselves in order to login. After logging in, the user can view movie details updated by the admin. In review page, the user can either add or view review for each movie. By viewing the review the user can choose the apt movie for their interest [11-16]. Then the added reviews will be stored in the EXCEL database, from where the reviews will categorize as true and fake reviews based on the text feature analysis. Then the positive and negative reviews will be used as the input for graph analysis for the purpose of graphical representation. The movie details will stored and fetched from the excel database which efficient for the data management. Figure 1 shows the Architecture design.

7. SYSTEM SOFTWARE

QT DESIGNER

PyQt5 is applied as a set of Python modules. It has over 620 classes and 6000 functions and technics. It is a multiplatform toolkit which goes on all major operating systems, including UNIX, Windows, and Mac OS. PyQt5 is dual licensed.

PyQt5 MODULES

Figure 1. Architecture design
PyQt5's classes are divided into several modules:

**QtCore**

The QtCore module contains the core non-GUI functionality. This module is used for working with time, documents and directories, various data types, and streams, URLs, mime types, threads or processes.

**QtMultimedia**

The QtMultimedia contains classes to handle multimedia content and Application Program Interfaces to access camera and radio functionality.

**QtNetwork**

These classes facilitate the coding of TCP/IP and UDP clients and servers by making the network programming easy and more portable.

**LANGUAGE USED – PYTHON**

![Python](https://example.com/python.png)

**Figure 2. Python**

Python is a general-purpose, dynamic, object-oriented programming language. The use of the Python language highlights coder efficiency and program readability. Python was first released in 1991. It uses indentation instead. Python is supported by a huge crowd of people throughout the world. Python is open source software. Python is a great initial point for people who need to learn coding as shown in Figure 2.

Python programming language supports several programming styles. Python supports object-oriented and procedural programming. There is also a restricted provision for practical coding.

8. HARDWARE COMPONENTS

**I5 PROCESSOR**

Intel's Core i5 processors tend to be more powerful than their i5 counterpart. Part of this comes down to fast asleep average clock speeds detected. Part of this comes down to additional cores of the processor.

**Features of I5**

I5 processors are used for intelligent communication and improve performance such as play, graphics etc. The I5 processor automatically manages the power supply where needed. I5 I have never violated the operation of the system. Allow the user to use a heavy-duty system with HD video composing, composing music and much more. I5 processors are able to increase system memory. Help users apply its best in
bandwidth and presentation. The ability to run it with two processors to work together is called two processors and can speed up efficient operation. Allows users to use their normal and important function with the help of heavy applications. The i5 processor contains Hyper Threading technology that enables users to perform multiple tasks and improve their business or work environment by performing two different tasks simultaneously.

9. RESULTS

The user interface to update movie details and viewing the reviews updated by the admin and viewing the movie details and updating reviews by the user/reviewer is designed successfully. The graph for the values of true and fake reviews is plotted successfully. Thus our system will help the user to view the reviews and the respective graph for those reviews which is an effective representation of good and bad opinions and thereby helps to find the correct opinions for their desired movie as shown in Figures 3-6.

![Figure 3. Main Window](image1)

![Figure 4. Review page](image2)
Fake reviews given by some of the negative persons would spoil the scope of the film industry by portraying a good movie as bad or bad movie as good one. So the customers would confuse and they themselves select something else, which may end up in huge loss for the film industry. Thus it is necessary to find a difference between true and fake review by detecting the false ones. But it is much more time consuming process if it is done manually by reading each and every review in order to detect the false reviews. Therefore, we proposed a technique to detect and classify the true and fake reviews in order to separate the fake reviews from the original ones. It works autonomously without human intervention. It uses Sentiment Analysis and text feature analysis to estimate the nature of review as good or bad. And also the proposed system will represent the categorization of reviews as true and fake reviews in a graphical manner which is user-friendly. The results shows that the proposed technique successfully predicted the fake reviews based on text emotions and categorized those reviews and also successfully performed the graphical analysis of those categorized reviews for each movie which is being updated in the platform. This will help in the development of film industry in good manner.
REFERENCES

[1] Streitfeld, D. Fake reviews, real problem. New York Times. http://query.nytimes.com/gst/fullpage.html, 2012.

[2] Rayana, S.; Akoglu, L. Collective opinion spam detection: bridging review networks and metadata. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 985-994.

[3] Catal, C.; Guldan, S. Product review management software based on multiple classifiers. IET Software, 2017, 11(3):89-92.

[4] Zuo, L.; Carass, A; Han, S.; Prince, J.L. Automatic outlier detection using hidden Markov model for cerebellar lobule segmentation. Proceedings of International Conference on Medical Applications in Molecular, Structural, and Functional Imaging, 2018, pp. 105780D-1-105780D-8

[5] J. Malbon, Taking fake online consumer reviews seriously, Journal of Consumer Policy, vol. 36, no. 2, 2013, pp. 139–157.

[6] R. Xia, C. Zong, and S. Li, Ensemble of feature sets and classification algorithms for sentiment classification, Information Sciences, vol. 181, no. 6, 2011, pp. 1138–1152.

[7] T. Barbu, Svm-based human cell detection technique using histograms of oriented gradients, cell, vol. 4, 2012, p. 11.

[8] G. Esposito, LP-type methods for Optimal Transductive Support Vector Machines. Gennaro Esposito, PhD, 2014, vol. 3.

[9] P. Kalaivani and K. L. Shummanathan, Sentiment classification of movie reviews by supervised machine learning approaches, Indian Journal of Computer Science and Engineering, vol. 4, no. 4, pp. 285-292, 2013.

[10] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225-5097-6.ch010

[11] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. 27, no. 4, p. 273, 2018.

[12] C.-H. Chu, C.-A. Wang, Y.-C. Chang, Y.-W. Wu, Y.-L. Hsieh, and W.-L. Hsu, Sentiment analysis on chinese movie review with distributed keyword vector representation, in Technologies and Applications of Artificial Intelligence (TAAI), 2016 Conference on. IEEE, 2016, pp. 84–89.

[13] V. Singh, R. Piryani, A. Uddin, and P. Waila, Sentiment analysis of movie reviews and blog posts, in Advance Computing Conference (IACC), 2013 IEEE 3rd International. IEEE, 2013, pp. 893–898.

[14] G. Vinodhini and R. Chandrasekaran, Sentiment analysis and opinion mining: a survey, International Journal, vol. 2, no. 6, 2012, pp. 282–292.

[15] G. Xu, Y. Cao, Y. Zhang, G. Zhang, X. Li, and Z. Feng, Trm: Computing reputation score by mining reviews, in AAAI Workshop: Incentives and Trust in Electronic Communities, 2016.

[16] N. Tian, Y. Xu, Y. Li, A. Abdel-Hafez, and A. Josang, Generating product feature hierarchy from product reviews, in International Conference on Web Information Systems and Technologies. Springer, 2014, pp.264–278.