



Movie Rating System Based on Sentiment Analysis

Rakshanda Mulay¹, Shivangi Pandey², Niranjana Patil³, Mehul Patil⁴, and Mrs. Seema Mandlik⁵

School of Computer Engineering and Technology

¹rsmulay@mitaoe.ac.in, ²slpandey@mitaoe.ac.in, ³niranjana.patil@mitaoe.ac.in, ⁴mdpatil@mitaoe.ac.in,

⁵sgmandlik@mitaoe.ac.in

MIT Academy of Engineering, Alandi(D).

Abstract

Manual reading of contents of movie reviews is a tedious and time consuming task for the movie viewers. So they are not able to make proper opinions/judgment to watch movies or not. Hence, this is our purpose to develop an automated movie rating system based on sentiment analysis. We have two sentiments which we are showing using emojis- Happy for positive sentiments and Sad for negative sentiments.

Keywords: Sentiment, Sentiment Analysis, Movie reviews, Human behavior, TextBlob

1. Introduction

In the last decade, there is a rise of social media and social networking sites like book my show, IMDB, film.com for movie reviews. It is a hectic, tedious, challenging and time consuming task for a movie viewer to read all the movie reviews which in turn results in inappropriate judgement/opinion about movies. Also instead of referring the newspaper that only contain a five star based rating which are usually not accurate, therefore, we propose an online system that manages the movie reviews given to the movie and predicts the rating corresponding to each movie review. It will show the overall rating of the movie which will change as the movie viewers enters the movie review. Emotions of movie viewers are displayed after analyzing the movie review given by him. Best and Worst reviews for particular movies are shown. Director, Writer, Movie stars, short summary of movie is shown for movie viewers' knowledge. The system also sorts and displays top rating movies as per analysis and calculate a top ten list automatically. This provides an automated movie rating system based on sentiment analysis which would help the movie viewers to make better choices of movie to

watch.

A. Intended Audience and Reading Suggestions

1. Novice Movie Viewers, such as a person who wants to know how well a movie has been received by the audience at large without going through each and every comment.
2. Researchers who would like to know about the software.
3. Tester who would like to know about the scope of software and must know about assumption.
4. Marketing staff would like to know about software special features.
5. Developers who would like to take the reference of the software.

B. Theory and Relevance

TextBlob library is used in the system for the sentimental analysis. TextBlob is a Python (supported for 2 and 3) library for text processing. It is a combination of multiple libraries and functions like NLTK, Google translator etc. to serve features like:

Features:

1. Part-of-speech tagging

2. Noun phrase extraction
3. Sentiment analysis
4. Classification (by using Naive Bayes, Decision Tree)
5. Language translation and detection powered by Google Translate
6. Spelling correction
7. Tokenization (splitting text into words and sentences)
8. Word and phrase frequencies
9. Parsing
10. n-grams

Word inflection (pluralization and singularization) and lemmatization

2. Related Work

The research papers used as reference for developing Movie rating system are as follows:

Kudakwashe Zvarevashe, Oludayo O Olugbara [1], proposed work involves design framework for sentiment analysis with opinion mining for Hotel customer reviews. Approach of sentiment analysis they used is detecting polarity considering of hotels facility with their respective aspects such as food and internet access of instance means they consider what people like or dislike about hotels services by supervised learning algorithm means the hotel features are labeled training dataset on hotel reviews that helpful to rate hotel. NLTK, OpenNLP and Stanford Core NLP are widely used for Basic NLP tasks. Success of sentiment analysis largely depends upon which tool is used for executing NLP tasks. They used Python with TextBlob (a library for processing textual data) gives better results and faster than NLTK and also provided a very easy interface for beginner to learn basic NLP tasks. They Built Intuition Model in which customers share their feelings and hotel reviews are automatically rated. Positive and negative feedback used for classifications. Classification is done by Naïve Bayes multinomial algorithm. But, we used only TextBlob library in our system not Naïve Byes multinomial algorithm because this library helps for data preprocessing and feature extraction is done by using Bags of words and classification is done by automatically based on sentiment polarity property which returned by

TextBlobAsiri Wijesinghe [2] addresses the problem that most sentiment analysis algorithms use simple terms to express sentiment about a product or service. However, cultural factors, sentence negation, sarcasm, terseness, language ambiguity and differing contexts make it extremely difficult to turn a string of written text into a simple pro or con sentiment. Sagar Chavan, Akash Morwal, Shivam Patanwala, Prachi Janrao [4] we refer only data preprocessing phase of this proposed system.

Our system and proposed system in paper of ZainabMirza, Mehwash Khan, Saima Khan, Khurshid Khatri [3] is almost same only difference is that they use of Naïve Bayes Algorithm decides the polarity of the comments with the help of which expert comments are provided. The unique thing of that system is that it extracts the user comments and automatically generates expert comments But, our system has special feature as compared to that proposed system in the paper of Zainab Mirza, Mehwash Khan, Saima Khan, Khurshid Khatri

[3] is we display sentiments of movie reviews using two emoji like Happy for positive sentiments and sad for Negative sentiments and show the overall rating of the movie which will change as the movie viewers enters the movie review and last important feature is we display result in graphical form like bar graph, pie chart for better analysis of reviews and help those movie viewers which have not seen yet that movie to take a better or quick decision of watch that movie or not ?

For better movie rating system our proposed system will satisfy following objectives:

1. To develop Interactive GUI those manages movie reviews and predict rating corresponding to each movie review.
2. To display the emotions of movie viewers after analyzing movie review given by him.
3. To show an average rating score of movie.
4. To analyze movie viewers review in graphical form.

3. Proposed System Architecture

The system architecture of proposed system is as shown in the above figure 1. The project is divided into four modules: Data Collection, Data Pre-Processing, Data classification and Data Analysis. In first module of the system, Data collection is done by collecting movie

reviews entered by the movie viewers in the form of English language only which get stored in the database and length of review is 5 words only.

In second module of the system, we pre-process data by removing white spaces, punctuations and other irrelevant words for feature extraction or keyword extraction. In third module of the system, data is classified into groups of sentiments like positive and negative on basis of sentiment polarity. In last module of the system, we analyze the movie review by generating different graphs.

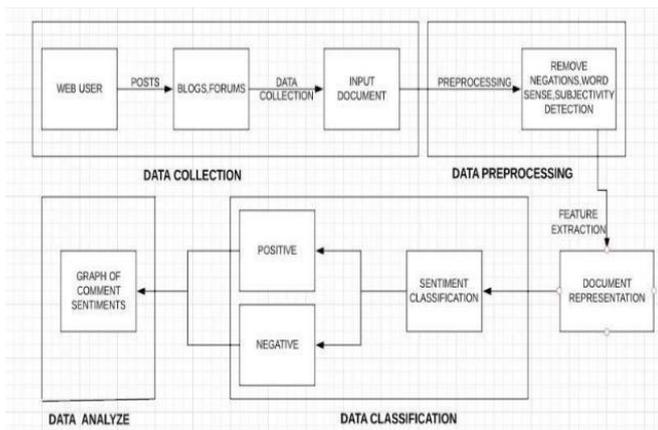


Figure 1: System Architecture

4. Methodology

Sentiment analysis on set of movie reviews which is given by movie viewers which measures the attitude of the movie viewers towards the movie i.e. whether it is positive or negative or try to understand what the overall reaction to the movie was according to them, i.e. if they liked the movie or they hated it. So, following modules are necessary for sentiment analysis on set of movie reviews.

A. Data Pre-processing

Incomplete and noisy data are common properties of real world database of movie review. Movie review needs to be cleaned before it is processed for better efficiency, accurate output of sentiment, to reduce the noise in review data and improve the overall classification performances. Movie review has been cleaned up somewhat, for example:

1. The movie review is in only English language.
2. All the text has been converted to lowercase.

3. There is white space around punctuation like periods, commas, & brackets.
4. Using an exhaustive stop word list (a, an, the, to for, as, in, with, etc.), some words can be ignored and removed during the pre-processing.
5. Review should not contain URLs.
6. Replace #HASTAG with only the word by removing the HASH (#) symbol. (i.e. #happy -> happy)
7. Replace emoticons with either EMO_POS or EMO_NEG.
8. Remove the vowels repeated in sequence at least three times, because by doing so the words are normalized: For example, two words written in a different way (i.e. coooooool and cool) will become equals.
9. Text has been split into one sentence per line.
10. Using correction function of TextBlob we do spelling correction.
11. We make use of the external python library PyEnchant, which provides a set of functions for the detection and correction of misspelled words using a dictionary.
12. Normalizing the dataset:

Words are often used with different variations. If these words are reduced to their root words, it would result in efficient performance of the algorithm.

Lemmatization and Stemming are techniques that are commonly used for normalizing the dataset. For example, words:

- Smile
- Smiling
- Laugh
- Laughing

All the above words can be reduced to the root word “smile”.

13. Eliminating features with very poor frequency:

Keywords that occur in lesser frequency usually do not play a role in text classification. User can get rid of these, resulting in better accuracy.

All the operations in this module are executed to try to make the text of movie review uniform. This is important because during the classification process, features are chosen only when they exceed a certain frequency in the movie review database. Therefore, after the basic preprocessing operations, having different words written in the same way helps the classification.

B. Feature Extraction

According to movie viewers, generally movie review consists of description of their thoughts about acting of actors, direction of movie, and story of movie, music of movie and overall experience of movie while watching.

Movie viewers mostly express their thoughts, emotions or opinion about movie in the form of special keywords or hash tags like (Poor Direction, Good Acting, Block Buster, Jhakkas, Inspiring story, Great Music, Refreshing, One time watch, Entertaining, Horror, Average, Bored Movie etc.) in movie review. These words are very helpful to predict sentiments or emotions of movie viewers. Hence we required this process.

Feature Extraction or keyword extraction is the process of finding special more important words or keyword which clearly expressed the emotions (happy, sad, anger or enjoy) of movie viewers about movie after removal of irrelevant, weakly relevant or redundant attributes or dimensions in a given movie reviews of movie.

C. Classification Phase

After feature extraction, Sentiment function of TextBlob returns two properties, polarity and subjectivity. Polarity is float value which lies in the range of $[-1,1]$ where 1 means positive movie review and -1 means a negative movie review. On basis of sentiment polarity of movie review our system program classifies movie review in two categories like positive and negative. However, we can

add more classes like neutral, highly positive, highly negative, etc.

D. Analysis Phase

Reviews can be analyzed by graphs like bar, pie or word cloud. This is very important phase because this phase gives final actual result for movie viewers.

5. Result Analysis

Social media contains a huge amount of sentiment data in the form of tweets, blogs, and updates on the status, posts, etc. Sentiment analysis of this largely generated data is very useful to express the opinion of the mass. As of now, there exists no such deployed platform that provides real time sentiment analysis on movie reviews of movie viewers. There are various systems wherein movie viewers can add their reviews or read other viewers' reviews of particular movie. The ratings are given by the movie viewers himself / herself. However, in our system, we are going to generate ratings based on movie reviews of movie viewers. A big advantage of TextBlob is, it is easy to learn and offers a lot of features like sentiment analysis, pos-tagging, noun phrase extraction, etc.

6. Conclusion

Sentiment based movie rating system aims at developing a website that manages the reviews given to the movie and predicts the rating corresponding to each review. It will show the overall rating of the movie which will change as soon as the movie viewers enters a new review. Our proposed System will have two types of users, registered user or unregistered user. Unregistered user can view the trailer of movie and can only read movie review of movie viewers. The registered user can view trailers give review about movie and see emotion of the movie viewers is displayed after analyzing the review given by him. Best and Worst reviews for particular movies are shown. Genre, Director, Writer, Running- time, Movie stars, short summary of movie is shown for movie viewers' knowledge. The system also display top rating movies as per analysis

In future work, we would like to work on Hindi and Marathi language movie reviews of movie viewers. Also take input of movie reviews of movie viewers in form of

video, audio and image and in our system registered movie viewers will get suggestions of movies based on his search history and the movie he/ she searches are visible in recently viewed section.

References

- [1]. Kudakwashe Zvarevashe, Oludayo O Olugbara, “A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews”, Conference on Information Communications Technology and Society (ICTAS), 2018, IEEE.
- [2]. Asiri Wijesinghe, “Sentiment Analysis on Movie Reviews”, Australian National University Technical Report – RESEARCH GATE, October 2015.
- [3]. Zainab Mirza, Mehwash Khan, Saima Khan, Khurshid Khatri, “Movie Rating System Based On Opinion Mining”, International Journal of Interdisciplinary Research and Innovations, Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015.
- [4]. Sagar Chavan, Akash Morwal, Shivam Patanwala, Prachi Janrao, “Sentiment Analysis of Movie Rating System”, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN: 2278-8727 , PP 43-47, 2017.
- [5]. Ramez Elmasri and Shamkant B. Navathe, “Fundamentals of Database Systems”, 7th Edition, 2017.
- [6]. Ramakrishnan and Gehrke, “Database Management Systems”, 3rd Edition, 2014.