



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Analysis of Movie Reviews with Prediction of Movie Characters

Saba Shireen Khan¹, Rajni Pamnani²

M. Tech Student, Dept. of Computer Engineering. K.J Somaiya College of Engineering, Vidyavihar, Mumbai, India

Asst. Professor, Dept. of Computer Engineering. K.J Somaiya College of Engineering, Vidyavihar, Mumbai, India

ABSTRACT: Sentiment Analysis offers opportunity to analyze user's emotions which is useful for decision making in many fields. One such field is Movie Industry, where sentiments of viewers affect on the success of any movie. In this thesis phrase-level Sentiment Analysis also known as Aspect based Sentiment Analysis is presented for analyzing various aspects of Movies from user reviews. This will help in predicting the best suitable characters for future movie projects. The aim is to summarize the Movie reviews at aspect level so that user can easily find out which aspects of Movie are liked or disliked by other users and calculation of overall rating based on aspects discussed in the reviews. Also based on user opinion the features of various actors will be extracted to predict the best suitable actors for future Movie projects. The algorithms proposed for the purpose are: Naive Bayesian classifier using supervised learning approach for classification of aspect polarity, Sentence parsing for finding the requirement of actors in the future projects & KNN (K nearest neighbor) algorithm for prediction of characters.

KEYWORDS: Sentiment Analysis; Naive Bayesian classifier; Aspect Polarity; Sentence Parsing

I. INTRODUCTION

Social media is emerging rapidly on the internet. This media knowledge helps people, company and organizations to analyze information for important decision making. Opinions which are being expressed in the form of reviews provide an opportunity for new explorations to find collective likes and dislikes of cyber community. One such domain of reviews is the domain of Movie reviews which affects everyone from audience, film critics to the production company. The Movie reviews being posted on the websites are not formal reviews but are rather very informal and are unstructured form of grammar. Opinions expressed in Movie reviews give a very true reflection of the emotion that is being conveyed. The presence of such a great use of sentiment words to express the review inspired us to devise an approach to classify the polarity of the Movie using these sentiment words.

Sentiment Analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment Analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service [5]. Users express their opinions about products or services they consume in blog posts, shopping sites, or review sites. It is useful for both the consumers as well as for the producers to know what general public think about a particular product or service. There are three levels in Sentiment Analysis, Document Level, Sentence Level and Aspect Level. Document-level and Sentence level opinions cannot provide detailed information for decision making. To obtain such information, a finer level of granularity is needed. Hence, the proposed method focused on Aspect based Sentiment Analysis, to obtain the information about movie and its characters. This analysis will be used for predicting the actors for future movie projects.

II. RELATED WORK

In [7] Authors have proposed a linguistic approach for Sentiment Analysis of message posts on discussion boards. For each sentence in a message post, dependency tree is generated, and splits the sentence into clauses. Then the contextual sentiment score for each clause is determined, by utilizing grammatical dependencies of words and the prior sentiment scores of the words derived from SentiWordNet and domain specific lexicons. In [8] the reviews of the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

Movies from different sources were collected and pre-processed to make it suitable for applying in the method. These pre-processed reviews are then passed through the aspect based text separator and the separated review text was obtained. The various Movie aspects used are screenplay, music, acting, plot, Movie, and direction. The functionality of the separator was to separate the review aspect wise. The separator used an aspect specific lexicon for the purpose of text separation. These aspect based separated sentences were given as an input to the classifiers meant for each aspects. A Naive Bayesian classifier was used for this purpose. It calculates the probability of a word or albeit a sentence, belonging to positive or a negative class of reviews. The traditional method of training and testing the classifier is applied. Based on the weightage of the driving factors of the Movie, the aspect based output is multiplied with the respective driving factor. Thus they have used a threshold score for the classification of the document.

In [9] Author focused on Genre Specific Aspect Based Sentiment Analysis of Movie Reviews. Using the aforementioned dataset and considering Movie genres like action, comedy, crime, drama and horror, they developed a fine grained unsupervised analysis model using lexicons that are context specific to each genre under consideration. The dataset taken was in XML format and each file contained Movie details like name of the Movie, genre of the Movie, date of release and full reviews. Then the aspect specific text was separated. Then sentiments orientation was found out by lexicon approach. In [10] before finding aspect and its respective opinion of Movie, proposed system performs subjectivity analysis. Subjectivity analysis is one of the important and useful tasks in Sentiment Analysis .Online reviews may consist of both objective and subjective sentences .Among these, objective sentences consist of only factual information and no sentiments or opinion. Hence subjective sentences are considered for further processing i.e. to find feature- opinion pair and to find summery at aspect level. In this paper, two different methods are implemented for finding subjectivity of sentences and then rule based system is used to find feature-opinion pair and finally the orientation of extracted opinion is revealed using two different method. Initially the proposed system uses SentiWordNet approach to find out orientation of extracted opinion and then it uses the method which is based on lexicon consisting list of positive and negative words.

In [11] the author examined the sentiment expression to classify the polarity of the Movie review on a scale of 0(highly disliked) to 4(highly liked) and performed feature extraction and ranking and use these features to train multilabel classifier to classify the Movie review into its correct label. In addition, a comparative study on different classification approaches has been performed to determine the most suitable classifier to suit the problem domain.

III. PROPOSED SYSTEM

A. Introduction:

Sentiment Analysis of a Movie review plays an important role in understanding the sentiment conveyed by the user towards the Movie. In the thesis work we focus on aspect based Sentiment Analysis of Movie reviews in order to find out the aspects of Movie liked or disliked by users. This analysis is also will be used to predict the actors for the future Movie projects based on the users liking to actors in certain roles.

B. Architectural Overview:

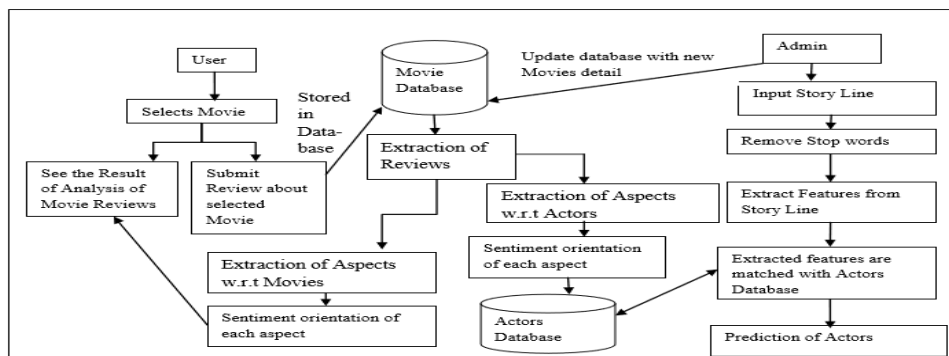


Figure 1 System Architecture

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

The architectural overview for our working model of the proposed system is shown in fig1. User can store their Reviews about particular Movie. These reviews will be used for analysis purpose. From the analysis, features discussed regarding the actors will be stored which will help in prediction of Movie characters. On submission of Story Line system will extract the Features of story hence Features of characters in the story. These features will be matched by stored parameters to predict the best suitable actor for the desired Role.

IV. METHODOLOGY & ALGORITHMS

Proposed Work is majorly divided into three parts:

1. Analysis of Movie Reviews
2. Feature Extraction from Story Line
3. Prediction of Characters

1. Analysis of Movie Reviews:

The proposed system uses customer reviews to extract aspect and mine whether given is positive or negative opinion. Each review is split into individual sentences. A review sentence is given as input to data preprocessing. Next, it extracts aspect in each review sentence. Stop word removal and stemming are data preprocessing. Sentiment orientation is used to identify whether it is positive or negative opinion sentence. Then it identifies the number of positive and negative opinions of each aspect.

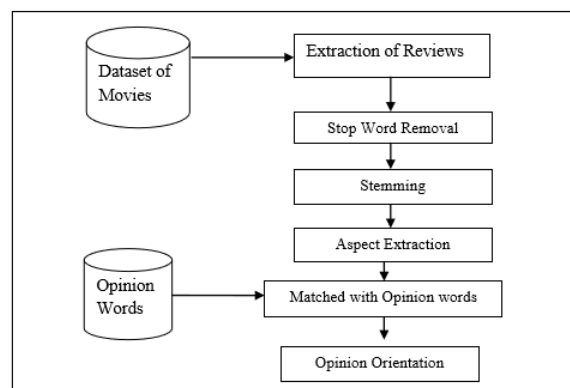


Figure 2 Aspect based sentiment analysis Process

- **Stop Word Removal:** Most frequently used words in English are not useful in text mining. Such words are called stop words. Stop words are language specific functional words which carry no information. It may be of types such as pronouns, prepositions, conjunctions. Stop word removal is used to remove unwanted words in each review sentence. Words like is, are, was etc. Reviews are stored in text file which is given as input to stop word removal. Stop words are collected and stored in a text file. Stop word is removed by checking against stop words list.
- **Stemming:** Stemming is used to form root word of a word. A stemming algorithm reduces the words "longing", "longed", and "longer" to the root word, "long". It consist many algorithms like n-gram analysis, Affix stemmers and Lemmatization algorithms. Porter stemmer algorithm is used to form root word for given input reviews and store it in text file.
- **Aspect Extraction:** This algorithm first extracts noun and noun phrases in each review sentence and store it in a text file. Aspects are extracted like story, action, dialogues, etc. Then, the frequent aspects are extracted and stored in text file.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

- **Sentiment Orientation:** Sentiment orientation will be found out using lexicon approach, and then the probabilities of positive and negative count are found according to the words using Naïve Bayesian classifier. The Algorithm is discussed below.

Algorithm:

String s1= "Extracted Opinion Word for which we need to find orientation"

Integer POS, NEG

Integer FLAG=0

String Orientation

1. If PosWd.contains (s1) Then FLAG=1; POS++;
Orientation is POSITIVE //Where PosWd is list of positive words.
2. If NegWd.contains (s1) Then FLAG=1; NEG++;
Orientation is NEGATIVE //Where NegWd is list of negative words.

After comparing all the words of the sentence, the found probabilities of the positive and negative counts are compared in the following manners.

- a) If the probability of positive count is greater than the negative count, then the sentence or opinion is positive.
- b) If the probability of negative count is greater than the positive count, then the sentence or opinion is negative.
- c) If the probability of positive count minus probability of negative count is zero, then it is neutral.

2. Feature Extraction from Story Line

Similar to Aspect Extraction Features from story Line is extracted. First story submitted in the database. All sentences will be extracted from story Line & on those sentences Sentence Parsing Algorithm will be applied. By doing so, we will extract information regarding different nouns available in the story and their adjectives i.e. requirements. This data will be stored in database.

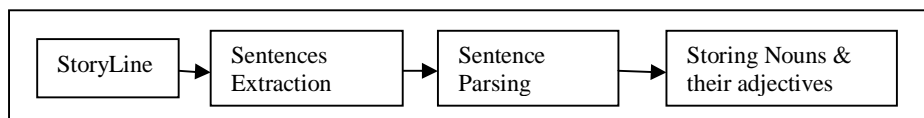


Figure 3 Feature Extraction Process

3. Prediction of Characters:

For prediction of character KNN (K nearest neighbor classifier) is proposed. KNN can be used to make clusters of actors having similar characteristics or having certain kind of characteristics. After requirements are extracted from storyline, it will be matched with KNN algorithm to find actors which are nearest to the role. KNN classifier learns by analogy, that is, it compares the given object with the training objects that are similar to it. The objects are defined by n attributes. KNN is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure. A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor [13].

The algorithmic steps are outlined below:

Input: The parameter k and a test object. The dataset, D

Output: The class label of the given test object

Algorithm:

- Compute similarity between the given test object and every other object in D.
- Select k objects from D those are closest to the given object.
- The output of the test object is the class label that occurs most number of times in the k selected objects.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 4, April 2017

V. CONCLUSION AND FUTURE WORK

By using Sentiment Analysis most important information will be extracted from Movie reviews. Such as information regarding various aspects of Movie will be extracted which convey the sentiments of user about aspects of Movies, i.e. whether it is liked or disliked by viewers. It will also give us insight of which actors are liked by viewer in which type of roles, resulting in contribution of viewers in the future Movie projects. This system further can be enhanced as a recommendation system, which will be used for recommending movies to users based on sentiment analysis results.

REFERENCES

1. A.Jeyapriya, C.S.Kanimozhi Selvi , "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm", IEEE Sponsored 2nd International Conference on Electronics and communication systems, pp 548-552,2015.
2. Bing Liu, 'Sentiment Analysis and Opinion Mining', Synthesis Lectures on Human Language Technologies, Morgan and Claypool Publishers, 2012.
3. Naveen Kumar Laskari, Suresh Kumar Sanampudi, "Aspect Based Sentiment Analysis Survey", IOSR Journal of Computer Engineering, Volume 18, Issue 2, pp 24-28, 2016.
4. Venkata Rajeev P, Smrithi Rekha V, "Recommending Products to Customers using Opinion Mining of Online Product Reviews and Features", IEEE sponsored International Conference on Circuit, Power and Computing Technologies, 2015.
5. https://en.wikipedia.org/wiki/Sentiment_analysis.
6. Amit Gupte, Sourabh Joshi, Pratik Gadgul, Akshay Kadam, "Comparative Study of Classification Algorithms used in Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 , pp 6261-6264 , 2014.
7. Tun Thura Thet, Jin-Cheon Na, Christopher S.G. Khoo, Subbaraj Shakthikumar, "Sentiment Analysis of Movie Reviews on Discussion Boards using a Linguistic Approach", ACM conference,pp 81-84,2010.
8. Viraj Parkhe, Bhaskar Biswas, "Aspect Based Sentiment Analysis of Movie Reviews", IEEE International Conference on Soft Computing and Machine Intelligence, pp 28-32, 2014.
9. Viraj Parkhe, Bhaskar Biswas, "Genre Specific Aspect Based Sentiment Analysis of Movie Reviews", IEEE sponsored International Conference on Advances in Computing, Communications and Informatics,pp 2418-2422, 2015.
10. Purtata Bhoir, Shilpa Kolte, "Sentiment Analysis of Movie Reviews Using Lexicon Approach", IEEE sponsored International Conference on Computational Intelligence and Computing Research, 2015.
11. Zbigniew Michalewicz ~ Martin Schmidt Matthew Michalewicz ~ Constantin Chiriac, " Adaptive Business Intelligence".
12. <http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms>.
13. Naganna Chetty, Kunwar singh vaisla, Nagamma Patil, "An Improved Method for Disease Prediction using Fuzzy Approach", IEEE sponsored Second International Conference on Advances in Computing and Communication Engineering, pp 568-572, 2015.
14. Dibya Jyoti Bora, Dr. Anil Kumar Gupta," A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm", International Journal of Computer Trends and Technology, Vol 10, pp 108-113, 2014.