International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 16 (2018) pp. 12788-12791
© Research India Publications. http://www.ripublication.com

Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques

Gurshobit Singh Brar [1,*], Asst. Prof. Ankit Sharma[2]

[1]*Department of Computer Science, Baba Farid College of Engineering and Technology, Bathinda, India. (*Corresponding Author)*

[2]*Department of Computer Science, Baba Farid College of Engineering and Technology, Bathinda, India.*

Abstract

Sentiment Analysis is a new subject in Research and is useful in many other fields. In Modern World, A huge amount of textual data is collected using surveys, comments, and reviews over the web. All of the collected data is used to improve products and services provided by both private organizations and governments around the world. This Paper includes sentiment analysis of movie reviews using feature-based opinion mining and supervised machine learning. In this paper, the main focus is to determine the polarity of reviews using nouns, verbs, and adjectives as opinion words. Reviews will be Classified into two different categories positive and negative. Reviews of Open Movie Database is used as source data set and Natural Language Processing Toolkit for Part of Speech Tagging. This paper also contains some facts about the classification of data on basis of polarity.

Keywords: Opinion Mining, Sentiment Analysis, Natural Language Processing, Sentiment Score, Sentiment Lexicon

INTRODUCTION

Every Human Being makes its decisions based on past experience, sentiments or opinion passed by other human beings. Whenever an individual wants to buy a new item or product, they seek opinions from others about the item or product. Similarly, every organization wants to deliver their best product to the market so they gather opinion from their customers about their product using surveys. Sentiment Analysis is a study of someone's opinions, sentiments or emotions expressed about a product or a movie.

A. Features Based Opinion Mining

In Features based Opinion Mining, relevant sentences are selected from a huge amount of data collected from surveys, comments, and reviews [26]. After extraction of useful data from a large chunk of data, keywords related to product features are extracted. There are steps for feature-based opinion mining.

1) Identifying Features

Identification of features is an important part of feature-based opinion mining. For Example, In the Sentence "The Movie has Great Storyline and Amazing Visual Effects." The Selected Features are "Storyline" and "VFX or Visual Effects". Hu and Liu [2] used a noun, nouns phrases and other parts of speech to identify features of a review.

2) Identifying Opinion Orientation

Determining the polarity (Positive, Negative or Neutral) of opinion is also a very important task. For Example, "This movie has Great Storyline and Amazing Visual Effects" is of positive polarity because both opinion words are of positive polarity. M. Annett, G. Kondrak [21] implemented a lexicon-based technique in "A comparison of sentiment analysis techniques: Polarizing movie Blogs". Lexicon based approach uses opinion words polarity to determine opinion Orientation and this work uses Lexicon and Naïve Bayes techniques for determining orientation and classification of opinions. There are other Techniques to identify Opinion polarity. Pang Lee [3] Used Naïve Bayes and Support Vector Machine for text classification.

3) Grouping of Synonyms

Many words in natural languages which have a similar meaning. For example, the word "Fierce" also means "Violent"," Vicious" and "Ferocious". In this step, we create a group of similar words together. In this work, a list of opinion words is already included therefore this part of feature-based mining is not used.

B. Model of Sentiment Analysis

Opinions are generally expressed for anything. For Example, a service, a product, a person, a topic, or an organization. The entity under observation has different components and may also have sub-components. Thus, the entity is called an object for sentiment analysis. Feature-based sentiment analysis uses the hierarchical model because objects are hierarchical in nature.

The object may have sub-components and attributes. Therefore, it is difficult for general people to understand these technical terms (attribute or components). So, a simple word "Feature" is used for featured-based opining mining. Opinion or sentiment can be expressed in one sentence or in multiple sentences as a paragraph. Opinion word orientation determines the orientation of opinion. One single sentence can one or more opinion words.

RELATED WORK

Minhoe Hur et al., 2016 [8] proposed a system to predict Box-office collection based on Sentiments of movie review. They have used Viewer opinions are used as input variables in addition to predictors and three machine learning-based algorithms (artificial neural network, regression tree, and support vector regression) were used to get non-linear

relationship between the box-office and its collection predictors.

Aurangzeb Khan, 2011 [19] proposed a rule-based technique in which SentiWordNet is used to obtain more accuracy than a pure lexicon-based technique for sentiment analysis for customer reviews and software reviews. The proposed system has 91% of accuracy at the document level and 86% of accuracy at the sentence level.

Mudinas and Zhang, 2012 [22] proposed a hybrid technique which gives better performance than the lexicon and almost performs like leaning based technique. Hybrid Techniques are stable as lexicon technique and performance as machine learning based techniques. The system has an overall accuracy of 82.3%.

Lei Zhang et al., 2010 [18] proposed a ranking and extracting product features in opinion documents algorithm. Initially, they have reviews of users and it was difficult to determine by the machine to differentiate between positive reviews and negative reviews. They used the associated rule mining technique for extracting product features.

Seven Rill et al., 2014 [13] proposed an application "PoliTwi" which shows Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis. In this Paper twitter, hashtags are used to determine the results of the election in the USA even before "Google Trends". Twitter API is used to collect data and Analyze using sentiment analyzing an algorithm.

Monu Kumar and Dr. Manju Bala, 2016 [7] proposed that it is difficult to analyze the huge amount of unstructured data these days gathered from various social networking sites like "Facebook, Twitter, and Instagram". Therefore, they used cloud service and utilized Hadoop for intelligent analysis and storage of big data. Sentimental Analysis of Twitter is done using the cloud.

Martin Wöllmer et al., 2013 [16] proposed a technique to analyze sentiments in Audio – Video context of a YouTube Movie. They used Metacritic database to get user reviews as input. They evaluated the knowledge-based approach, applying data-based approach in an in-domain setting as well as in a cross-domain setting.

Giuseppe Di Fabbrizio et al. 2013 [15] proposed aspect rating distributions and language modeling which used for summarizing online product and service reviews. They used a novel approach for extracting multi-document summarization for textual data that considers aspect rating distributions and language modeling as summarization features.

Rafeeque Pandarachalil et al., 2014 [12] proposed a method for Twitter sentiment analysis using an unsupervised learning approach. They determined the Polarity of tweets is evaluated by using three sentiment lexicons-SenticNet, SentiWordNet, and SentislangNet. They used parallel python framework to implement this method.

Chirag Sangani 2013 [14] proposed a method for analyzing user sentiments towards apps through their review comments and ratings can be economically profitable to app developers. They propose a system that provides a list of reviews for each topic that represents user opinions towards that topic and a many-to-many relation portraying from reviews to topics of interest.

PROPOSED WORK

A. Data Input

There are two ways to give input to the movie review sentiment analyzer. One by providing a list of reviews in JSON file format. Or by providing the TMDB ID of Movie Title.

In Case if TMDB ID a TMDB [27] JSON API is used to fetch and store reviews in MySQL Database. After fetching reviews, first 10 reviews for a particular title are used by the system for sentiment analysis.

B. Part of Speech Tagging

POS is used to disambiguate a sentence in order to extract features from a sentence [2]. In POS tagging each word is labeled. It is used to determine word position in the grammatical context. POS tagging helps to find out nouns, noun phrases, verbs and adjectives in a sentence. After POS Tagging there is a little chance selected word is a discarded word for feature selection and opinion words.

C. Features and Opinion Words Extraction

All opinion words are selected from the sentence. The system extracts all nouns, noun phrases, verbs and adjectives from the movie review and compares with the existing list of words. These words are classified on basis of their polarity. For Example "good" word is of positive polarity. On the other hand, features are selected on basis of number times occurrence of opinion words. If opinion word is an occurrence in review higher than the threshold value then it is added features list. For this system API is trained only for movie reviews with keyword and phrases dictionary which includes "good acting", "solid story" and "awesome action".

D. Identify Sentence Polarity

After extracting all features and Opinion words, it is very easy to find the polarity of the sentence. Sentence polarity follows the same rules as arithmetic expressions. A negative sentiment contains all negative opinion words and positive sentiment contain all positive opinion words. A negative sentiment may contain a positive opinion word. For Example: "This movie Story is not good" sentence in a movie review. In this sentence, "good" opinion word is of positive polarity but "not" is a negative word. Therefore, the overall polarity of this sentence will be negative.

E. Identify Review Polarity

Whole review polarity depends on a number of total positive or negative sentences found in a review. If the number of total positive sentences is greater than the number of total negative sentences then review polarity will be positive. Similarly, a review polarity will be negative if the number of total negative sentences is greater than the number of total positive sentences.

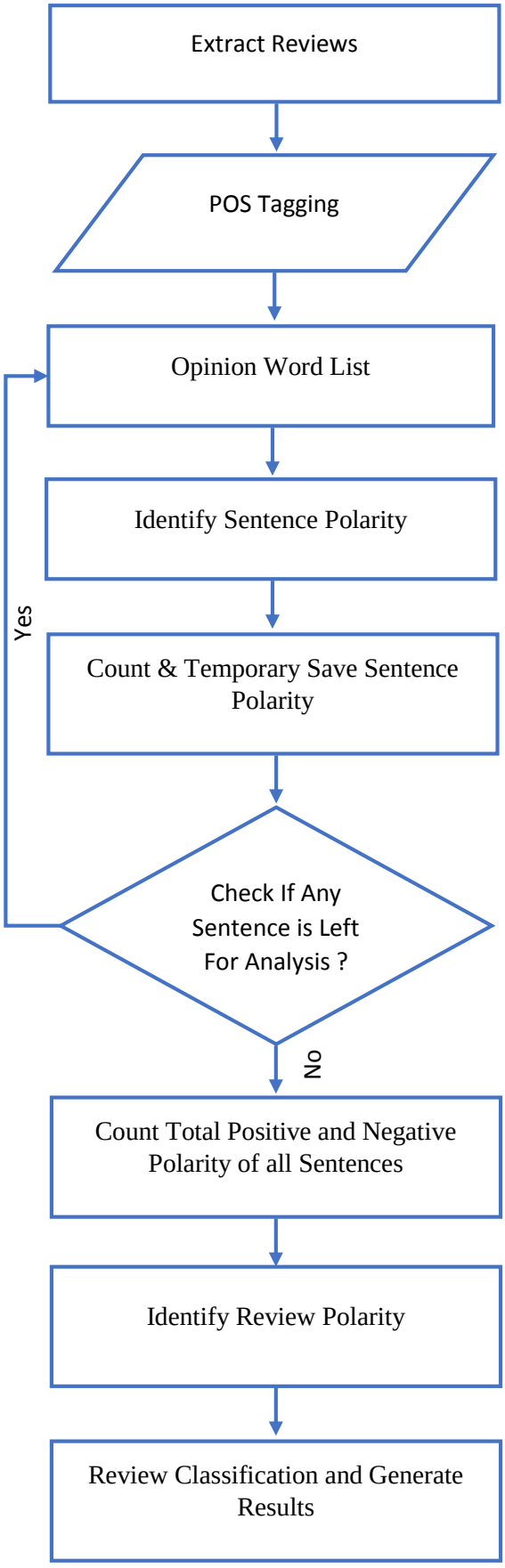

Figure 1: Flowchart of Proposed System

F. Classification of Review

Once, review polarity is calculated. Review polarity percentage and polarity (Positive or Negative) classified [28] and saved for further analysis. With further analysis, box office collection can be predicted and overall performance of movie can also be predicted.

EXPERIMENTAL EVALUATION AND RESULTS

The system is tested for 50 plus different movie titles each with max 10 reviews and final results of total 500 reviews are shown in table 1. It shows no of provided features, the accuracy of the system (Accurate Result Percentage), Error Percentage and False Negative Percentage and False Positive Percentage. False Negative means a positive polarity review considered as negative. Similarly, False Positive means a negative polarity review considered as positive. The average accuracy of this system for test review is 81.22%.

Table 1. Results of Proposed System

Total No. of Titles	60
Total No. of Reviews	500
Avg. Accuracy (%)	81.22%
Avg. False Positive (%)	8.14%
Avg. False Negative (%)	11.08%

CONCLUSION

In this paper, movie reviews are classified into positive or negative polarity. The system proposed by author in the paper can be used to classify a huge database of movie reviews. Best thing about the system that it is a web-based API for sentiment analysis for movie reviews with JSON output to display results on any operating system. Table 1 shows that the system works decently. This will help movie producers to check the status of their movie. Future work, this API can be trained for other reviews like smartphones, laptops or clothes etc.

REFERENCES

[1] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–135.

[2] M. Hu and B. Liu, "Mining and summarizing customer reviews," Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2004, pp. 168-177.

[3] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques,” Proceedings of the ACL-02 conference on Empirical methods in natural language processing, vol.10, 2002, pp. 79-86.

[4] Jie Yang University of Wollongong, Australia “Mining Chinese social media UGC- a big-data framework for

analyzing Douban movie reviews", Journal of Big Data Springer, 2016

[5] Kia Dashtipour Scotland, United Kingdom "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques", Springer, 2016

[6] Kigon Lyu Korea University, Korea "Sentiment Analysis Using Word Polarity of Social Media", Springer, 2016

[7] Monu Kumar Thapar University, Patiala "Analyzing Twitter sentiments through big data", IEEE, 2016

[8] Minhoe Hur Seoul National University "Box-office forecasting based on sentiments of movie reviews and Independent subspace method", Information Sciences, 2016

[9] Jorge A Balazs University of Chile "Opinion Mining and Information Fusion- A survey", 2015

[10] Donglin Cao Xiamen University, China "A cross-media public sentiment analysis system for microblog", Springer, 2014

[11] Min Chen Huazhong University of Science, China "Big Data: A Survey", Springer, 2014

[12] Rafeeque Pandarachalil Govt. College of Engineering, Kannur "Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach", Springer, 2014

[13] Seven Rill Goethe University Frankfurt, Germany "PoliTwi- Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis", Elsevier, 2014

[14] Chirag Sangani Stanford University, USA "Sentiment Analysis of App Store Reviews", 2013

[15] Giuseppe Di Fabbrizio A&T Research Labs, USA "Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling", Digital Object Identifier IEEE, 2013

[16] Martin Wöllmer Technical University of Munich, Germany "YouTube movie reviews- Sentiment analysis in an audio-visual", IEEE Computer Society, 2013

[17] Andranik Tumasjan Technical University of Munich, Germany "Predicting elections with Twitter - what 140 characters reveal about political sentiment", 4th International AAAI Conference, 2010

[18] Lei Zhang University of Illinois, Chicago "Extracting and Ranking Product Features" Coling 2010: Poster Volume, pages 1462–1470, Beijing, 2010

[19] A. Khan, B. Baharudin, K. Khan; "Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure" ICSECS 2011: 2nd International Conference on Software Engineering and Computer Systems, Springer, pp. 317-331, 2011.

[20] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," Proceedings of WWW, 2003, pp. 519–528.

[21] M. Annett, G. Kondrak, "A comparison of sentiment analysis techniques: Polarizing movie Blogs", In Canadian Conference on AI, pp. 25–35,2008.

[22] A. Mudinas, D. Zhang, M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, pp. 1-8, 2012.

[23] H. Wang, Yue Lu, and C. Zhai. Latent aspect rating analysis on review text data: a rating regression approach. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 783-792. ACM, 2010.

[24] S. Moghaddam and Martin Ester. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In Proceedings of the 19th ACM international conference on Information and knowledge management, pages 1825-1828. ACM, 2010.

[25] W. Zhang, H. Xu, W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects-based sentiment analysis," Expert Systems with Applications, Elsevier, vol. 39, 2012, pp. 10283-10291

[26] M. Chen, Huazhong University of Science, China "Big Data: A Survey", Springer, 2014.

[27] https://api.themoviedb.org/3/ - Movie Reviews Input API

[28] https://www.nltk.org/ - POS Tagging and Classification

[29] https://github.com/japerk/nltk-trainer - NLTK Trainer