

Evaluation of Television Shows Popularity Based on Twitter Data using Sentiment Examination Techniques

D. N. V. S. L. S. Indira, JNVR Swarup Kumar, Y. Adi Lakshmi, N. Rajeswari

Abstract: A Blogging platform promotes you to conferrer expose your current status in the type of short posts by making utilization of texting, social media along with online messages. In the meantime, this is constantly producing large amounts of unsorted and un-organized information which has turned out to be a perplexing responsibility for examination. Twitter, the interpersonal organization benefit, is a rich wellspring of data on client reaction to an occasion, for example, a TV program. Here, we present a contextual investigation where we consider freely accessible tweets that were posted when a famous TV indicate was publicized. We utilize standard content mining methods to dissect the tweets and know the rating of show. This analysis can be used by producers for future shows for improving their business.

Index Terms: Popularity, Television, Twitter Data, Sentiment, Analysis.

I. INTRODUCTION

The exact and auspicious forecast of program notoriety is of incredible incentive for substance suppliers, publicists, and communicated TV manager. Twitter is the place customers post small messages called tweets to their supporters and perused tweets of people whom they seek after. Twitter is the place clients post little messages called tweets to their supporters and examined tweets of individuals whom they look for after. The Twitter API gives authorities of twitter applications the gadgets and systems to empty twitter information. The twitter API starting at now has three portions - a REST API that gives client data, a Search API that offers access to the chase capacities of twitter and finally a Streaming API that gives advancing access to high volume reviewed and filtered data [1]. This paper thought about twitter for information gathering. . In this paper, an open source procedure for content mining and opinion investigation using a game plan of R groups for mining Twitter data and conclusion examination is shown [2]. In this paper, we considered three different categories of television shows. The shows are Big Boss, CID and Game of Thrones. We gather tweets with respect to the shows. Applied content mining and assessment investigation on the tweets and rate the show. SVM classifier is used for mining and scoring messages.

Revised Manuscript Received on September 20, 2019.

D. N. V. S. L. S. Indira*, Department of IT, Gudlavalluru Engineering College, Gudlavalluru, INDIA, indiragamani@gmail.com.

JNVR Swarup Kumar, Department of CSE, Gudlavalluru Engineering College, Gudlavalluru, INDIA, swarupjnvr@yahoo.co.in

Y. Adi Lakshmi, Department of CSE, Gudlavalluru Engineering College, Gudlavalluru, INDIA, laxmi072003@gmail.com

N. Rajeswari, Department of CSE, Gudlavalluru Engineering College, INDIA, rajeswari.gec@gmail.com

II. RELATED WORK

Twitter has been utilized for sentiment analysis in numerous examinations [5, 6] for different prospects[2].For example, Zhou et al. furthermore, Tumasjan et.al [6], came up with a technique for mining feelings from twitter about dominant races applicants and foreseeing the decision results. Regardless, existing guess procedures as a rule require a significant measure of tests and long getting ready time and the desire accuracy is poor for tasks that experience a high zenith or sharp decrease in pervasiveness Supposition investigation can be done using two ways. The foremost one is feeling vocabulary based philosophy, in which, the lexicon is built out of a course of action of great as well as troublesome end words, utilized to score the appraisal sentences either, ideal, adverse or fair-minded [7]. This system is astoundingly outstanding and needs an enumerating ability to score every sentence as demonstrated by the nearness of great or troublesome terms [2].

The next procedure accepts machine learning frameworks for setting up a classifier adopting a game plan of pre-describes appraisals as a readiness set. By then, use the readied categorizer to describe new emotions as ideal, troublesome or unbiased sets [2].

III. PROPOSED WORK

An algorithm has been utilized to investigate the general assessment depending upon on tweets of twitter. This includes:

1. Twitter Data extraction to CSV file
2. Pre-Processing of Data
3. Sentiment Scoring
4. SVM classifier

A. Twitter Data extraction to CSV file

RStudio is programming used to associate the client with the twitter application. With the ultimate objective to isolate tweets from twitter, we use twitter Rest API to interface with Twitter API. Tweets escaped in the CSV record.

- A twitter account is created.
- An application is created using user developing tools in the twitter
- Tweets are received using Hashtag
- Trending TV shows Hashtags are given to extract data.
- Pre-Processing stage utilizes these documents for extricating the fundamental information

B. Pre-Processing of Data

There few pre-processing steps to clean the text data

1. Replace blank spaces
 Replace blank spaces with data
2. Replace @username
 Replace @username with only username
3. Remove punctuation
 Remove punctuations in the tweets
4. Remove links
 Remove the links in the tweets
5. Remove tabs
 Remove tab spaces in the tweets
6. Remove blank spaces at beginning
 Remove blank spaces in the beginning of tweets
7. Remove blank spaces at ending
 Remove blank spaces in the ending of tweets
8. Remove digits
 Remove digits in the tweets
9. Convert text to lower case
 Converts the whole text into lower case
10. Corpus creation

A corpus is created for the text and to remove stop words [3] in the text.

A word cloud is created to the corpus. Frequent terms can be represented in a graph to know how frequent the words appear in the text.

C. Sentiment Scoring

- A group of favourable and unfavourable words are gathered and saved in a text file format.
- Based on these text documents the tweets text is scored.
- The text is split into words.
- Compare the words in the text document and the words of tweets.
- If the words match in positive set then it is positively rated.
- If the words match in negative set then it is negatively rated.
- Otherwise it is considered as neutral word.
- Score is calculated by considering the difference between all the positive and negative number of words.
- A data frame is created to the scores and return scores.

D. SVM Classifier

Support Vector Machine (SVM) is a portrayal and backslide figure device that adopts speculation of machine learning to extend perceptive accuracy during normally maintaining a vital division from over-fit to the data [4]. Be that as it may, it is for the most part utilized in order issues. Support vector machines can be characterized as systems which avail concept of straight space limits in a high dimensional component space, organized with proceedings in computation from enhancement speculation that performs a proceeding in inclination obtained from perceptible grasping speculation [4].

Steps to implement SVM classifier:

- Read the file which has sentiment score and mined text in it into a variable.
- Create a matrix for text, which will be used for training data.

- Read a file which has only mined text in it into a variable.
- Create a matrix for text, which will be test data.
- Create containers for text with sentiment score and test data.
- Now train the model using the first container.
- Apply SVM for both train set and test set.
- Save the results in CSV file.
- Plot the results.

IV. EXPERIMENT RESULTS



Fig 1: Word cloud for bigboss



Fig 2: Word cloud for CID



Fig 3: Word cloud for game of thrones



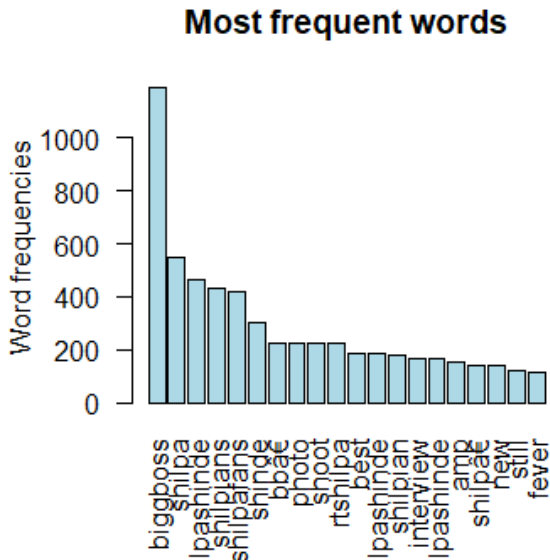


Fig 4: Frequency graph for bigboss

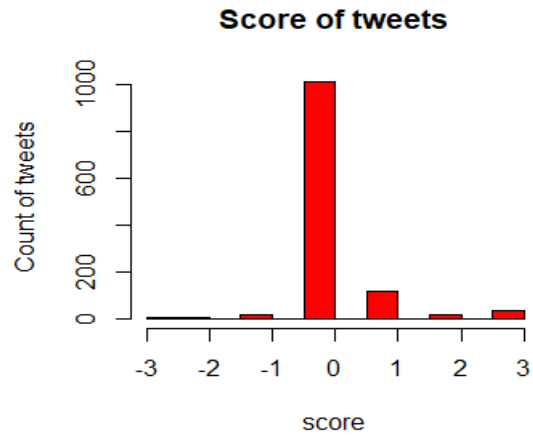


Fig 7: Sentiment analysis of bigboss histogram

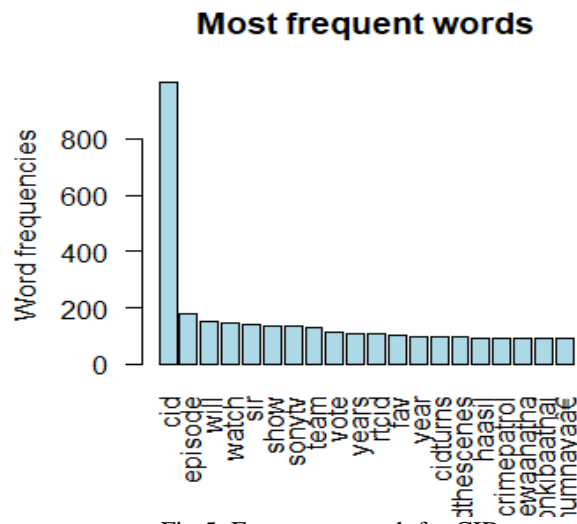


Fig 5: Frequency graph for CID

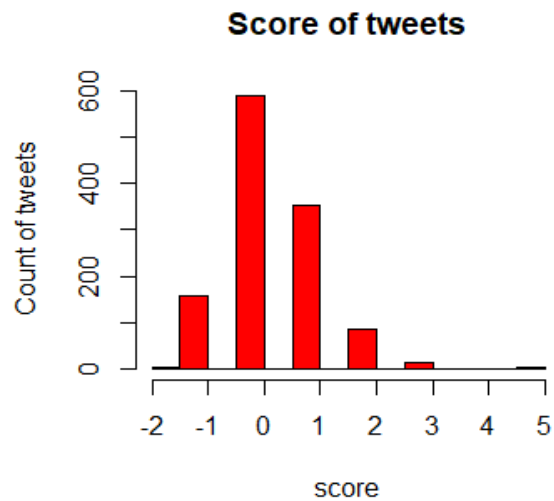


Fig 8 : Sentiment analysis of game of thrones histogram

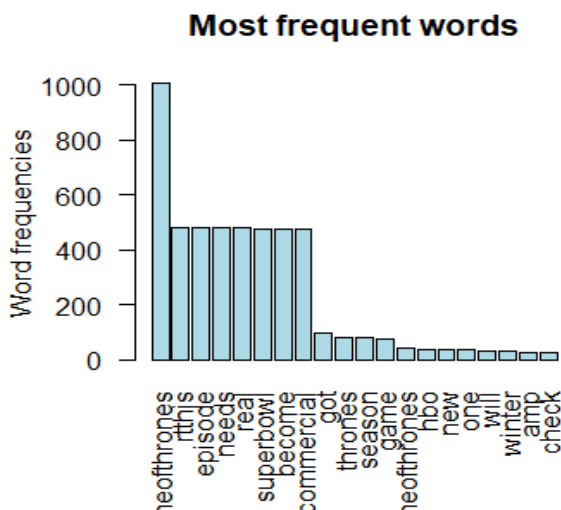


Fig 6: Frequency graph for game of thrones

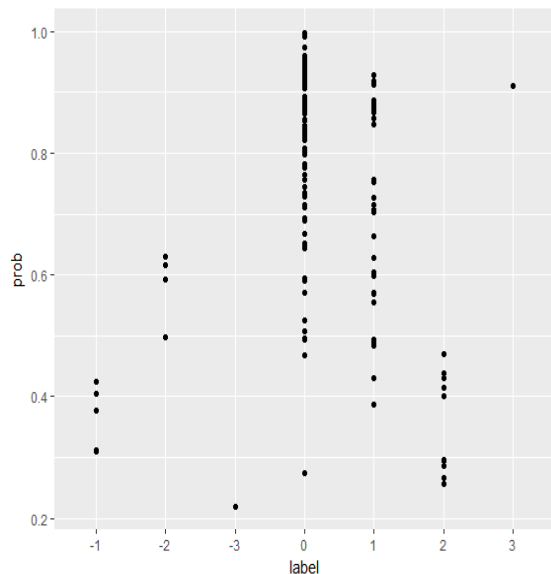


Fig 9: SVM Classification for bigboss

Rating	Negative		Neutral	Positive			
	-2	-1	0	1	2	3	5
Sentiment Score							
Frequency	1	158	590	352	85	13	1

Table2: Showing Sentiment scores and their Frequency for cid data.

Rating	Negative				Neutral	Positive	
	-	-	-	-	0	1	2
Sentiment Score	5	3	2	1			
Frequency	1	8	7	7	947	13	27

Table3: Showing Sentiment scores and their Frequency for game of thrones data.

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have gathered the publicly available tweets using Rest API. The collected tweets are pre processed, analyzed and provided with a score. The mined text is classified and applied with SVM classifier. Based on the scores and rating, the producers uses business intelligence to produce better shows and improve the current shows.

The results also help the producers to produce sequential for high rated shows or re-telecast the highly rated show. The producers consider the drawbacks of the low rated shows and develop them to get high rates.

In this case study we have used only twitter data and it can be further extended to other social media data such as facebook, youtube etc.,. This case study we have considered a particular show in a particular language. It can be extended to find the rating of same show in different languages.

REFERENCES

1. Chengang Zhu , Guang Cheng , Kun Wang “Big Data Analytics for Program Popularity Prediction in Broadcast TV Industries”, IEEE, 2017
2. Eman M.G. Younis.Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study.2015
3. D.N.V.S.L.S.Indira, R.Kiran Kumar. Polarity Classification and Intensification based on Emoticons and POS Tagging on Twitter Data.
4. Vikramaditya Jakkula. “Tutorial on Support Vector Machine (SVM)”
5. Pak, A., &Paroubek, P. (2010, May). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In LREC (Vol. 10, pp. 1320-1326).
6. Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013, June). Sentiment analysis on tweets for social events. In Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on (pp. 557-562). IEEE.
7. Milstein, S., Lorica, B., Magoulas, R., Hochmuth, G., Chowdhury, A., & O'Reilly, T. (2008). Twitter and the micro-messaging revolution: Communication, connections, and immediacy--140 characters at a time.

AUTHORS PROFILE



Dr. DNVSLs Indira has received PhD from Krishna University and published 21 research papers in different International Conferences and International Journals. She is Peer - Reviewer for International Journals/Conferences and she is a life member of IE.



Mr. NVR Swarup Kumar J is pursuing PhD from Annamalai University and published number of books/research papers in different International Conferences and International Journals. He is Peer - Reviewer for International Journals/Conferences. He is Member of IEEE (91099158)



Mrs. Y Adilakshmi is pursuing PhD from JNTUK University and published 15 research papers in different International Conferences and International Journals and she is a life member of CSI.



Mrs. N.Rajeswari is pursuing PhD from JNTUK University and published 12 research papers in different International Conferences and International Journals and she is a member of ACM.

