JOURNAL OF MULTIDIMENSIONAL RESEARCH & REVIEW

Website: http://www.jmrr.org **Vol.2, Iss.2 Pages:25-31**

SENTIMENT ANALYSIS ON NEWS HEADLINES

[1] S Dhinesh Kumar, [2] S Omprakash

II Year M.Sc Decision and Computing Sciences, Coimbatore Institute of Technology, Coimbatore

Article Received: April 2021 Published: July 2021

Abstract

Sentiment Analysis is one of the emerging fields in Natural Language Processing. It is the process of detecting positive or negative sentiment in text. Newspaper is a main source of information that encompasses what is happening around the world to the majority of people in India. Hence by these headlines, we can understand the situation of the country. In this paper, we are going to analyse the news headlines published on Times of India using NLP techniques. The objective of this study is to investigate hundreds of thousands of news headlines and polarise them into three categories of positive, negative and neutral.

Keywords: Sentiment Analysis, Natural Language Processing, Polarity, Subjectivity, News Articles

I INTRODUCTION

We acquire the information in the outer world in the form of news. Nowadays news comes in different formats and variations. Since the last decade, the internet has made the news to spread swiftly and easily in which people can get updates on their mobile devices itself. Though there are numerous ways of updating ourselves with the current updates, the most traditional source which is the newspaper is still widely popular among the majority of people in India and the world. India in particular has the world's second largest newspaper market in the world behind China. As per reports in 2018, India had combined circulation of about 240 million copies. People have started to access the newspapers digitally in the wake of Internet facilities across India.

Sentiment Analysis is contextual mining of text which identifies and extracts subjective information in source material, and helps a business to understand the social sentiment of their brand, product or service while monitoring online conservations. However, analysis of social streams is usually restricted to just basic sentiment analysis and count based metrics. Traditional sentiment analysis involves using references dictionaries of how positive certain words are and then calculating the averages of these scores as the sentiment of that text.

The primary objective or goal of this study is to analyse hundreds of thousands of news headlines obtained from Times of India from the year 2001- 2020 in India and present the percentage of sentiment scores of the news articles obtained. Finally, analyse if people in the country are relatively happier or not.

The format of this paper is as follows: Section II presents the related work conducted in sentiment analysis. Section III elucidates the methodology of paper used in analysis. The outcomes or the results are explained on Section IV of paper. Section V explains the limitations of our research work. The conclusion is presented on Section VI.

II RELATED WORKS

There has been an increase in the number of researches undergone in the field of Sentiment Analysis in the last decade. Discussed the works of other researchers in this section.

Soonh Taj, Baby Bakhtwer Shaikh, Areej Fatemah Meghji have made two approaches for the Sentiment Analysis of News Articles. First, they have discussed a lexicon-based approach for sentiment analysis of the news articles. Lexicon based approach means that the words are assigned values for sentiment through training data. Secondly, they have used machine learning techniques such as Natural Language Toolkit (NLTK) and Parts of Speech (POS) syntactic models. This particular paper concluded that categories of business and sports had more positive articles whereas entertainment and tech had the majority of negative articles [1].

Prakashini S and Vijayakumar D have elucidated on the topic Sentimental Classification of News Headlines using Recurrent Neural Network. In this paper, they have proposed the application of Recurrent Neural Network with Long Short Term Memory Unit (LSTM). Eventually they have estimated the accuracy of the classification algorithm used in this paper [2].

Namrata Godbole, Manjunath Srinivasaih and Steven Skeina from Stony Brook University, USA have presented a system that assigns scores which indicates positive or negative opinion on the text corpus. They have also elucidated on how sentiment scores can vary by geographical location, demography and type of news source [3].

Arpit Upadhyay, Nishi Sharma, Aryan Chaudhary and Divya Jain have performed Sentiment analysis of generalized text and tweets [4]. They have explained the functioning of sentimental analysis and how to connect to Twitter and sentiment analysis queries. They then use the tweepy library to collect huge amounts of positive, negative and neutral tweets. They use NLP packages such as Textblob and spacy. In this paper, they are exploring Parts of Speech separate from unigram models and contain POS information within their unigram models in the future time.

III METHODOLOGY

In this study, approach to carry out is Sentiment analysis. Sentiment Analysis can be generally undergone by two approaches: supervised and unsupervised. Supervised approach has a bunch of training data that is used to build a classification model with the intent of using this model to classify new data for which labels or not present. Unsupervised approach to sentiment analysis does not require training data.

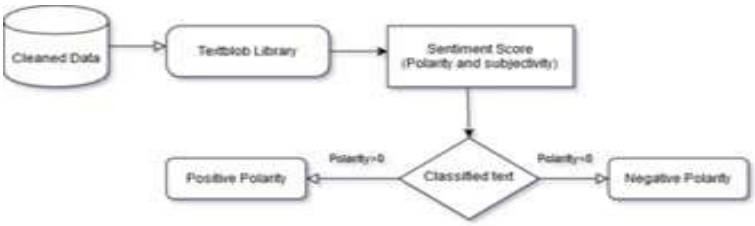

Fig 1: Working Methodology of our study

A. Dataset Collection:

The dataset used is extracted from Times of India which is India's largest news website. The dataset contains 3 columns and over 3 million rows which symbolizes the news texts. The columns are Date, category of headlines and texts of headlines.

B. Dataset Pre-processing:

Preprocessing is an essential step to perform analysis on the data. Preprocessing is nothing but weeding out the inconsistent and irrelevant data present in the huge dataset. For our dataset, we have used the process of "Tokenization" where we break a sequence of sentences into individual parts such as words, phrases and symbols. We have also got rid of the punctuation marks. After this process, we have converted the news text to the lower case using the "Transform Cases" operator.

C. Calculating Sentiment Scores of News Articles:

Sentiment Scores are usually calculated by finding polarities of each news text and sentences and finally merge them to find the overall polarity. Polarity in sentiment analysis refers to identifying sentiment orientation such as positive, negative or neutral. In general, language can contain expressions that are objective and subjective. Objective expressions are nothing more than facts whereas Subjective expressions are opinions that describe people's feelings towards a specific subject or topic.

For instance, consider the following expressions:

- Coimbatore is a developing city.
- The 2004 tsunami left many families with sorrow and grief

First expression represents a fact and the second expression represents many families' emotions.

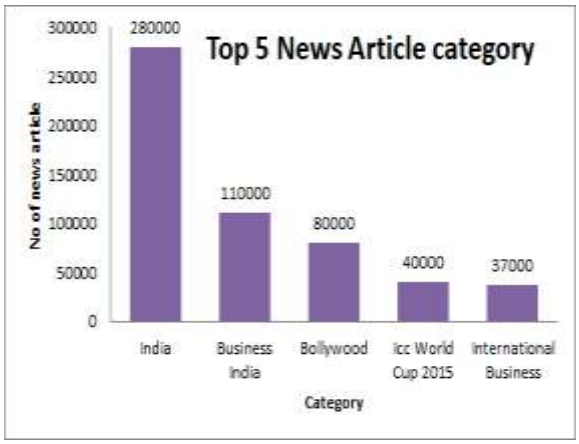

Fig 2 Top News Article Category

Fig 2 explains the categories of articles that are discussed in the news texts in our dataset. It is obvious that the topic of India is the most discussed article category followed by Business,

Bollywood, ICC world cup of 2015 and about international business. There are still numerous categories of articles but we have just jotted down the top 5 categories.

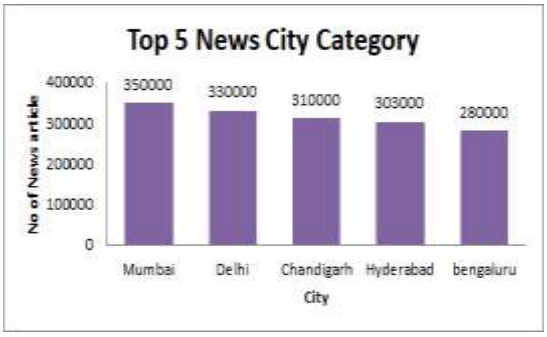

Fig 3 Top News City Category

Fig 3 interprets about the news articles in the city category in India obtained from our Times of India dataset. The cities that are in the top 5 categories are the metropolitan cities in our country since there are a huge chunk of people residing there. Hence it is clear that more people equal to more news stories. The cities in the top 5 categories are as follows: Mumbai, Delhi, Chandigarh, Hyderabad and Bengaluru. As discussed before, we have explained only the top 5 cities in this category.

IV RESULTS OBTAINED

The news texts are classified into three major classifications: positive, neutral and negative. The scores greater than 0 are treated as positive sentiment whereas the scores lesser than 0 are treated as negative sentiment and the news articles having sentiment score of 0 is considered as neutral.

Table 1 Percentage of News Sentiments

TYPE OF NEWS SENTIMENT	PERCENTAGE
Positive Headlines	19
Negative Headlines	12
Neutral Headlines	69

The percentages of the types of sentiments are tabulated clearly in Table 1. The percentages of positive headlines are about 19 whereas the percentages of negative headlines are about 12 and the neutral sentiments are the majority of news articles are about 69.

By these sentiment scores, we can clearly understand the situation of the country. Hence if we eliminate the neutral or unbiased headlines, we can comprehend that there are more positive news articles than negative news texts. News articles reflect the current affairs and problems currently happening in the country. Eventually if positive scores are greater than the negative scores, people in India are relatively happier.

V LIMITATIONS

We have only used the Times of India's news headlines from 2001 to 2020. But India has over 1,00,000 registered newspapers which are published in over 50+ languages. This is one of the most important drawbacks of our project. Sentiment Analysis techniques focus on the texts in English and Chinese only.

Our study uses only one source of news articles for the sentiment analysis performed. This cannot also predict the overall happiness of the people in India.

VI CONCLUSION

Numerous breakthroughs are applicable for the sentiment analysis that can be performed. This paper uses Sentiment Analysis on news articles obtained from Times of India on the years 2001-2020. This model classifies the total news articles into three broader sentiment categories which are positive, neutral and negative. Future work of this research could be implemented the concept of sentiment analysis on various other Indian languages.

REFERENCES

1. Soonh Taj. Baby Bakhtwer Shaikh and Areej Fatemah Meghji, Sentiment Analysis of News Articles: A Lexicon based Approach
2. Prakashini S and Vijayakumar D, Sentimental Classification of News Headlines using Recurrent Neural Network
3. Namrata Godbole, Manjunath Srinivasaiah and Steven Skeina, Large-Scale Sentiment Analysis for News and Blogs, In International Conference on Weblogs and social media, Denver, CO, 2007.
4. Arpit Upadhyay, Nishi Sharma, Aryan Chaudhary and Divya Jain, Sentiment Analysis of Generalized Text and Tweets
5. M. U. Islam, F. B. Ashraf, A. I. Abir and M. A. Mottalib, Polarity detection of online news articles based on sentence structure and dynamic dictionary, 2017 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, 2017, pp. 1-5. doi: 10.1109/ICCITECHN.2017.8281777
6. B. Meyer, M. Bikdash, and X. Dai, Fine-grained financial news sentiment analysis, SoutheastCon 2017, 2017.
7. M. Devika, C. Sunitha, and A. Ganesh, Sentiment analysis: a comparative study on different approaches, Procedia Computer Science, vol. 87, pp. 4449, 2016.
8. Dandrea, F. Ferri, P. Grifoni, and T. Guzzo, Approaches, tools and applications for sentiment analysis implementation, International Journal of Computer Applications, vol. 125, no. 3, pp. 2633, 2015.

Cite this article:

S Dhinesh Kumar, S Omprakash, "Sentiment Analysis on News Headlines", Journal of Multidimensional Research and Review (JMRR), Vol.2, Iss.2, pp.25-31, 2021.