International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 31st Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/71

Title COMMENT CLASSIFICATION USING SENTIMENT ANALYSIS

Volume 12, ISSUE 03, Pages: 510-516

Paper Authors
Prasanth Kumar.K, Hema Latha. N, Rushitha.Ch, Sai Prasanth Babu.D, Kalyan.K

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Comment Classification using Sentiment Analysis

Prasanth Kumar.K[1], Assistant Professor, Department of Information Technology,Dhanekula Institute of Engineering & Technology, Vijayawada.
Hema Latha. N[2], Graduation Student, Department of Information Technology, Dhanekula Institute of Engineering & Technology, Vijayawada.
Rushitha.Ch[3], Graduation Student, Department of Information Technology, Dhanekula Institute of Engineering & Technology, Vijayawada.
Sai Prasanth Babu.D[4], Graduation Student, Department of Information Technology Dhanekula Institute of Engineering & Technology, Vijayawada.
Kalyan.K[5], Graduation Student, Department of Information Technology, Dhanekula Institute of Engineering & Technology, Vijayawada.

Abstract
Taking into consideration, the use of social media platforms like Twitter, YouTube, Facebook, Ginger, as well as others is expanding. Everyone seems to be free, to express their opinions, whether they are stated in text or video-based. Anyone utilizing a social media network (precisely we are considering YouTube) may encounter multiple diversity comments. As a consequence, the individual may not examine every single comment to determine whether they are positive, neutral, or negative. It can be a lengthy process and gets hectic. The user can get to know about the community acceptance of its channel/video based, on that one can maintain their content quality. To overcome this, we are implementing neural networks to classify the comments using Sentiment Analysis. The results demonstrate that our system, which uses modern techniques, does well on the categorization test, demonstrating its ability to aid content producers boost audience engagement on their channel.

Keywords: Classification, Natural Language Processing, Sentiment Analysis.

INTRODUCTION:
In recent years, YouTube has become tremendously popular among individuals who create videos. Many content creators publish their video content on this platform. 60 hours of video are uploaded to YouTube every minute, or one hour each second[1]. These videos have received a lot of views and comments. The creators of the content, more typically referred to as YouTubers, must consistently strive to maintain the standard and pace of their output. They must accomplish this by seeking input from their viewers in the comment stream. This response allows them to comprehend the significance of their work. Feedback not only fosters audience involvement but also reveals content improvement opportunities. Not all YouTubers, though, have the time to look through every comment. The user can get to know about the community acceptance of its channel/video based, on that one can maintain their content quality Our research tries to address a solution to this problem. Our method entails extracting all of the comments from a video and grouping them based on sentiment analysis. These distinctions may contribute to the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

content fascination According to mood and sentence structure we extracted every remark from a video and divided them into a variety of categories. Positive, Neutral, and Negative are of these groupings. This pairing may be advantageous curiosity regarding their content. In this project, we scrape and categorize YouTube comments into several groups in accordance with their attitudes. We made the decision to broaden the scope of sentiment research in order to escape the bubble of just positive, negative, and neutral sentiment. There are benefits of sentiment analysis which includes Sorting Data at Scale, Real-Time Analysis and consistent criteria.

Recurrent Neural Networks

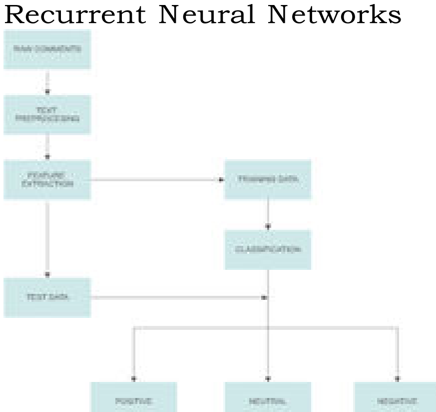

Fig.1. Process Diagram

Since they can handle sequential input, recurrent neural networks (RNNs) have been frequently employed in sentiment analysis. RNNs are frequently used in sentiment analysis to model the context of a text by taking into account the words that came before it in the sequence. The Long Short-Term Memory (LSTM) network is one of the well-liked RNN types utilized in sentiment research. The issue of vanishing gradients in conventional RNNs, which can make it challenging for the network to learn long-term relationships in the data, is resolved by LSTMs. In order to selectively recall or forget data from earlier timesteps, LSTMs employ a set of gates to regulate the flow of information through the network. Machine translation tasks have shown great promise for recurrent neural networks (RNNs). Unlike feedforward neural networks, RNNs can handle variable-length sequence[2] input because they have a recurrent hidden state whose activation at each iteration is dependent on the activation at the previous iteration. A typical RNN is seen in Figure 3. The illustration depicts the unfolding of an RNN into a complete network. The network would be unrolled into a 4-layer neural network, one layer for each word, for instance, if the input sequence was a four-word phrase. The following are the formulae that control an RNN's computations. Because the sentiment of a phrase may rely on the sentiment of earlier portions of the text, they are therefore ideally suited to representing sentiment in text. RNNs are often trained on labelled datasets where each text is assigned a sentiment label before being used for sentiment analysis (e.g. positive, negative, neutral). The sentiment label of fresh, unread messages is then predicted using the RNN[3]. RNNs have generally been proved to be successful in jobs involving sentiment analysis, producing cutting-edge outcomes on several benchmark datasets. They may, however, be computationally expensive and need a lot of data to train well.

A. Web Scraping

Using automated software or tools, web scraping is the process of collecting data from websites. For analysis or other purposes, the extracted data can subsequently be kept in an organized manner, such a spreadsheet or database. Online scraping may be used for a number

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

of things, including research, pricing comparison, data mining, and rival website monitoring[4].For web scraping, a variety of tools and methods are available, such as: Online scraping tools like Beautiful Soup, Scrapy, and Requests are available in programming languages like Python. extensions and plugins for the browser, including Data Miner, Web Scraper, and Scraper. instruments for online web scraping like Import.io, Octoparse, and ParseHub.

B. Sentiment Analysis

The Sentiment analysis, commonly referred to as opinion mining, is a method for obtaining subjective data from text or audio data[5]. It entails locating and categorizing the thoughts, feelings, and attitudes stated in a text, such as news stories, product evaluations, or social media posts. Finding out if a text's sentiment is favorable, negative, or neutral is the main objective of sentiment analysis. Many methods, such as rule- based systems, machine learning, and natural language processing, can be used for sentiment analysis (NLP). To better the consumer experience and create focused advertising campaigns, it is frequently utilized in marketing and advertising. Political analysis, social media monitoring, and customer service all make use of sentiment analysis.

C. Model evaluation

There are a number of indicators that may be used to gauge a sentiment analysis model's success.Accuracy, which counts the percentage of cases that were properly categorized out of all instances, is a frequently used metric[5]. Accuracy, however, can be deceiving if the data is unbalanced, that is, if one class (for example, positive or negative) is significantly more prevalent than the other. Other measures like accuracy, recall, and F1-score may be more instructive in such circumstances.Recall is the percentage of real positive (or negative) cases that were properly identified, whereas precision indicates the proportion of instances that were classified as positive (or negative) that are truly positive (or negative).for longer time; it utilizes smaller place and heats slower than signal phase motors.

LSTM

Recurrent neural network (RNN) architecture known as Long Short-Term Memory (LSTM) was developed to address the shortcomings of conventional RNNs in processing sequential input. For applications requiring long-term dependencies and memory, such language modelling, audio recognition, and sentiment analysis, LSTMs are very helpful. The utilization of memory cells and gates to regulate information flow across the network is the primary characteristic of LSTMs. Memory cells and gates control the information flow into and out of the memory cells, allowing LSTMs to keep a long-term memory of earlier inputs. The gates consist of:

The forget gate regulates how much data is erased from the memory cell[6].The input gate regulates how much data is added to the memory cell.Controls how much data is output from the memory cell using the output gate.

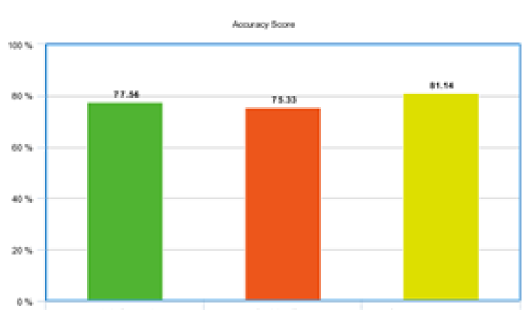

Fig.2. Comparison of accuracy score

The LSTM may learn how to selectively retain or forget information based on the input data since the gates are controlled by activation functions like sigmoid and tanh.

Prerequisites

To follow along the reader will need the following:

1. Python installed on your system or access to the Google Colab.
2. A Dataset available in the form of a CSV file.

Natural language processing (NLP) activities including text creation, sentiment analysis, and machine translation have all been demonstrated to benefit from the use of LSTMs[7]. They are frequently used in conjunction with other deep learning methods, such as language modelling and convolutional neural networks (CNNs) for text categorization.

RNN Techniques

1.Basic RNN: Often referred to as a vanilla RNN, this is the most basic form of RNN. The incoming data is processed one token at a time by a single layer of recurrent neurons[8]. The vanishing gradient issue, however, prevents it from handling long- term dependencies effectively.

2.Long Short-Term Memory (LSTM): LSTM is a sort of RNN architecture that controls the information flow across the network using memory cells and gates. It has been demonstrated to be efficient in sentiment analysis tasks and can successfully handle long-term dependencies.

3.Gated Recurrent Unit (GRU): GRU is a less complex LSTM variation that makes use of fewer gates and parameters. It has been demonstrated to be efficient in tasks involving speech recognition and language modelling.

4.Bidirectional RNN: A bidirectional RNN analyses the input data both forward and backward, enabling it to gather contextual data from both past and future tokens. For sentiment analysis jobs that call for a deeper comprehension of the input data, this may be helpful.

5.Attention-Based RNN: While creating predictions, an attention-based RNN focuses on particular areas of the input data using an attention mechanism. This is beneficial for sentiment analysis jobs when certain input data elements are more crucial than others.

Python implementation

As usual to any implementation, we get started with fetching the dataset and preparing it ready for our model implementation.

Step 1:- Web scraping, scrape all the the comments.

Step 2:- Input the data by custom scripting.

Step 3:-This plot will represents the different YouTube channle comments snetiments.

Long Short-Term Memory

Recurrent neural network (RNN) architecture known as LSTM (Long Short- Term Memory) is an excellent choice for sentiment analysis jobs[8]. Based on the general emotional tone of the text, sentiment analysis entails categorising content into categories of positive, negative, or neutral sentiment. LSTM networks are a potent technique for sentiment analysis because they can accurately simulate long-term relationships and sequential patterns in the text.

LSTM networks have been used in a variety of methods for sentiment analysis applications. One typical method is to employ an LSTM

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

network that has already been trained to extract features from the text input, which can then be fed into a classification model to predict sentiment. This strategy has shown to be successful for tasks like the sentiment analysis of movie reviews and social media posts[9]. An alternative strategy is to specifically tune an LSTM network for the given sentiment analysis job. In order to increase the accuracy of sentiment classification, the LSTM network's parameters must be optimised after training on a collection containing text data and sentiment labels. This strategy has been proven to work well for applications like sentiment analysis of consumer comments and product evaluations.

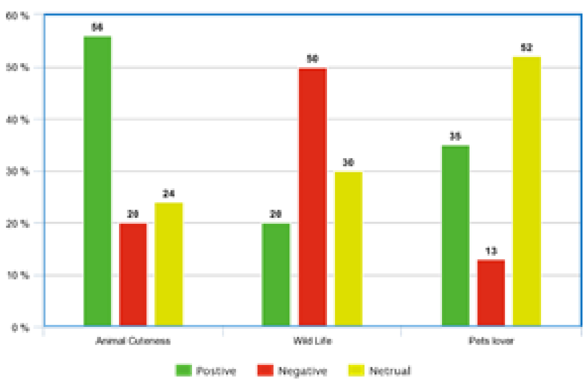

Fig.3 different channels comment responses

Several recurrent neural network designs, like GRU (Gated Recurrent Unit) and bi- directional RNNs (Bidirectional Recurrent Neural Networks), have been employed for sentiment analysis tasks in addition to LSTM, but LSTM continues to be one of the most well-liked and efficient choices. In general, LSTM networks are an effective tool for sentiment analysis tasks, particularly when working with lengthy text data sequences.

Algorithmic Steps

As for the model development and classification of the YouTube comments, the data is extracted and put together. The data is tried to classify into different sentiments they are, say Positive, Negative and Neutral.

1. Illustrates the step-by-step procedure starting from the very first step that is, Web Scraping and so on followed up by some computations and concluding with the classification[9].

2. Analyzes the percentage of comments that fall under positive, negative and neutral in a pie chart which is nothing but graphical representation.

3. Considering any YouTube channel say 'x', taking up a video comments and classifying them into positive, negative and neutral and comparing them, but putting them together in a bar graph representation.

By following up the entire procedure user can get to know about the community acceptance of its channel/video based, on that one can maintain their content quality[10].

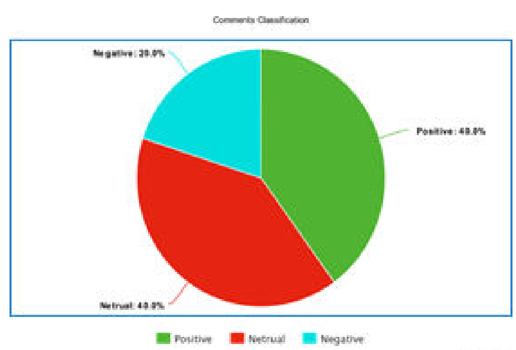

Fig.4. Percentage of sentiments

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

S.No	Algorithm	Accuiracy
1	Logistic Regression	77.56
2	Decision Tree	75.33
3	Recurrent neural network	81.14

Table-1. Classification Models Accuracy

The suggested techniques will provide the most accurate sentiment analysis of comments on YouTube. The percentages of accuracy provided by each model are displayed in Table above. Based on the accuracy of the predicted outcomes and the labelled data, will show it the graph between each YouTube channel. The final example would display the comment classification as a pie chart.

Conclusion

In this study we investigated the efficiency of the three machine learning algorithms to build classifiers that are precise and reliable. This includes the Random Forest (RF),Logistic Regression (LR), and Decision Tree algorithms. We have trained deep learning model using recurrent neural neural network for the classifiaction of the YouTube comments. It started with web scraping and is followed by the Pre-processing of the dataset. Pre-processing includes data labeling, lowercasing of the text, stop words removal, data splitting, and feature extraction. For the sentiment classification into three classes positive, negative and neutral. Recurrent neural network deep learning classification algorithm has been used and achieved a noticable accuracy on training data and on test data, which is obviously more reliable than previous.

References:

[1] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," Journal of Business Research, vol. 69, no. 2, pp.897–904, 2016.

[2] P. Russom et al., "Big data analytics," TDWI best practices report, fourth quarter, pp. 1–35,2011.

[3] S. Erevelles, N. Fukawa, and L. Swayne, "Big data Consumer analytics and the transformation of marketing," Journal of Business Research, vol. 69, no. 2, pp.897–904, 2016.

[4] V. Hatzivassiloglou and K. R. McKeown,"Predicting the semantic orientation of adjectives," in Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics.Association for Computational Linguistics,1997, pp. 174–181.

[5] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase- level sentiment analysis," in Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2005, pp. 347–354.

[6] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004, p. 271.

[7] A. Pak and P. Paroubek, "Twitter

as a corpus for sentiment analysis and opinion mining." in LREc, vol. 10, no. 2010, 2010.

[8] M. WAHYUDI and D. A. KRISTIYANTI, "Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization." Journal of Theoretical & Applied Information Technology, vol. 91, no. 1, 2016.

[9] D. N. Devi, C. K. Kumar, and S. Prasad, "A feature based approach for sentiment analysis by using support vector machine," in Advanced Computing (IACC), 2016 IEEE 6th International Conference on. IEEE, 2016, pp. 3–8.

[10] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced naive bayes model," in International Conference on Intelligent Data Engineering and Automated Learning. Springer, 2013, pp. 194–201.

[11]