International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Response analysis of Educational videos

Muskan Chowatia[1], Sayali Jadhav[2], Gargi Kelaskar[3], Nileema Pathak[4], Suchetadevi Gaikwad[5]

[1,2,3]Student, Information Technology Department, Atharva College of Engineering, Mumbai, India
[4,5]Asst. Professor, Information Technology Department, Atharva College of Engineering, Mumbai, India

---***---

Abstract - *In recent years online learning has become an integral part of the education system, the main reason being that online environment provides flexibility to access anytime and anywhere. Online learning portals like Coursera, YouTube, Udemy, etc. have gained the attention of many students by providing courses from a growing number of selected institutions as well as many instructors who upload their course lectures online. We are gathering and analyzing students' feedback about the teacher's performance, the learning experience, other course attributes, etc. The use of sentiment analysis of student feedback can furthermore, give an understanding of student response towards teaching approach like to determine students' interest in a class and to identify areas that could be improved through corrective actions and essentially determine classroom mood.*

***Key Words*: online learning, sentiment analysis, feedback, student response, data mining**

1. INTRODUCTION

Sentiment analysis is the process of extracting and analyzing text data for measuring sentiment polarity. Online learning videos on E-learning portals are becoming an effective training approach these days. Due to this, the students can interact and share doubts on certain topics. However, the teachers often remain outside of this process and cannot understand the learning problems that exist in their classrooms. Hence to get an in depth overview, we use sentiment analysis so that the teacher can better tune their teaching approach. Sentiment analysis of student feedback is a form of indirect assessment that analyzes textual response written by students whether in formal course surveys or informal comments from online platforms to determine students' interest in a class and to identify areas that could be improved through corrective actions.

1.1 Literature Survey

B Elizabeth Poch Nishant Jha, Grant Williams, Jazmine Staten, Miles Vesper, Anas Mahmoud had proposed a system to analyze User Comments on YouTube Coding Tutorial Videos. The main objective of this paper is to help content creators to effectively understand the needs and concerns of their viewers, thus respond faster to these concerns and deliver higher-quality content.

Sujata Rani and Parteek Kumar had proposed a system to analyze student comments from course surveys and online sources to identify sentiment polarity, the emotions expressed, and satisfaction versus dissatisfaction. A comparison with direct assessment results demonstrates the system's reliability.

Chei Sian Lee, Hamzah Osop, Dion Hoe-Lian Goh, Gani Kelni, had proposed a system where a total of 150 educational videos on YouTube were selected and 29,386 comments extracted using our customized extraction software application. Sentiment and qualitative content analyses were performed.

Nabeela Altrabsheh, Mohamed Medhat Gaber, Mihaela Cocea had proposed a system This paper will discuss how feedback can be collected via social media such as Twitter and how using sentiment analysis on educational data can help improve teaching. The paper also introduces our proposed system *Sentiment Analysis for Education* (SA-E).

2. PROPOSED SYSTEM AND IMPLEMENTATION

2.1 PROPOSED SYSTEM

Our system architecture has five main components: data collection, data preprocessing, sentiment and emotion identification, satisfaction and dissatisfaction computation, and data visualization. The system uses the open source python language to perform data preprocessing and sentiment classification.

The block diagram in Figure 1 shows in schematic form the steps involved in pre-processing of the data.

1. Lower casing

The first pre-processing step which we will do is transform our text into lower case. This avoids having multiple copies of the same words. Characters are converted to lower case to ease the process of matching words in student comments to words in the NRC Emotion Lexicon.

2. Removing Punctuation

The next step is to remove punctuation, as it doesn't add any extra information while treating text data. Therefore, removing all instances of it will help us reduce the size of the training data. All the punctuation, including '#' and '@', has been removed from the training data.

3. Removal of Stop Words

As we discussed earlier, stop words (or commonly occurring words) should be removed from the text data. For this purpose, we can either create a list of stopwords ourselves or we can use predefined libraries.

4. Common word removal

Previously, we just removed commonly occurring words in a general sense. We can also remove commonly occurring words from our text data First, check the 10 most frequently occurring words in our text data then take call to remove or retain.

5. Rare words removal

Similarly, just as we removed the most common words, this time we remove rarely occurring words from the text. Because they're so rare, the association between them and other words is dominated by noise. You can replace rare words with a more general form and then this will have higher counts

6. Spelling correction

There would definitely exist comments with a plethora of spelling mistakes. In that regard, spelling correction is a useful pre-processing step because this also will help us in reducing multiple copies of words. For example, "Analytics" and "analytcs" will be treated as different words even if they are used in the same sense. To achieve this we will use the *textblob* library.

7. Tokenization

Tokenization refers to dividing the text into a sequence of words or sentences. In our example, we have used the *textblob* library to first transform our tweets into a blob and then converted them into a series of words.

8. Stemming

Stemming refers to the removal of suffices, like "ing", "ly", "s", etc. by a simple rule-based approach. For this purpose, we will use *PorterStemmer* from the NLTK library. For example, *dysfunctional* has been transformed into *dysfunct*, among other changes.

9. Lemmatization

Lemmatization is also an effective option like stemming because it converts the word into its root word, rather than just stripping the suffices. It makes use of the vocabulary and does a morphological analysis to obtain the root word.

All these pre-processing steps are essential and help us in reducing our vocabulary clutter so that the features produced in the end are more effective.

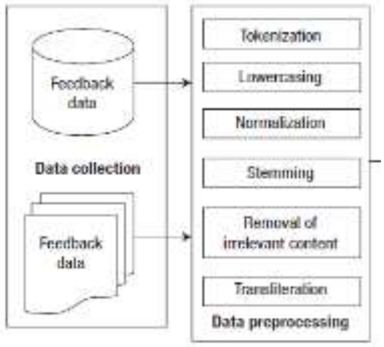

Fig-1 : Data pre-processing overview

2.1 IMPLEMENTATION

We used a combination of Lexical method and TF-IDF approach. RandomForest classifier was used to train our model. Term frequency is the ratio of the count of a word present in a sentence, to the length of the sentence.

Therefore, we can generalize term frequency as:

TF = (Number of times term T appears in the particular row) / (number of terms in that row).

The following diagram depicts the tf calculation of a sample text comment.

	words	tf
0	awesome	2
1	thing	1
2	wanted	1
3	told	1
4	knowso	1
5	short	1
6	video	1

Fig 2 : Term Frequency

3. Inverse Document Frequency

The intuition behind inverse document frequency (IDF) is that a word is not of much use to us if it's appearing in all the documents. Therefore, the IDF of each word is the log of the ratio of the total number of rows to the number of rows in which that word is present.

IDF = log(N/n), where, N is the total number of rows and n is the number of rows in which the word was present.

So, let's calculate IDF for the same tweets for which we calculated the term frequency.

The more the value of IDF, the more unique is the word.

4. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF is the multiplication of the TF and IDF which we calculated above.

	words	tf	idf
0	awesome	2	2.818398
1	thing	1	2.595255
2	wanted	1	4.204693
3	told	1	4.204693
4	knowso	1	4.204693
5	short	1	3.511545
6	video	1	1.806797

FIG 3 : IDF calculated for a sample

Data visualization of the sample set of the comments we extracted.

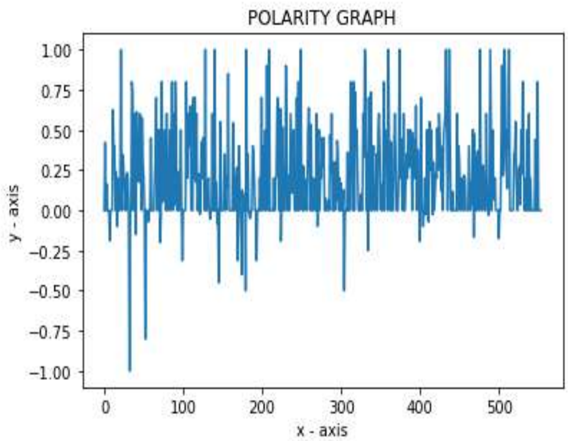

Fig 4 : Polarity Graph

The above line graphs shows the polarity of comments ranging from -1 to 1 (where -1 is negative,0 is neutral and 1 is positive).

e 'don't', 'can't', and 'use' because they are commonly occurring words. However, it has given a high weight to "wanted" since that will be very useful in determining the sentiment of the tweet.

2.2 VALIDATION

The following image depicts the evaluation of our model's performance using precision, recall and f1-score. Our classifier achieved an accuracy of 62.5 percent.

	precision	recall	f1-score
neutral	1.00	0.25	0.40
positive	0.57	1.00	0.73
accuracy			0.62
macro avg	0.79	0.62	0.56
weighted avg	0.79	0.62	0.56

0.625

FIG 5 : Accuracy achieved by system

3. CONCLUSION

Classification of general events of user comments in YouTube is a challenging task for researchers so far. A lot of work is done in this regard. Many research papers have emphasized on following problems in order to find the response of comments given by the users of YouTube. Current sentiment dictionaries having limitations, informal language styles used by users, estimation of opinions for community-created terms, to assign proper labels to events, to achieve satisfactory classification performance and the challenges involving social media sentiment analysis.

The evaluation of responses can highlight the topics, which are more complex for the students. In this way, at the end of the course teacher can improve the teaching approach.

REFERENCES

[1] Sujata Rani and Partook Kumar, "A Sentiment Analysis System to improve teaching and learning", IEEE Computer Society Journal 2017

[2] E. Poche and N. Jha, "Analyzing user comments on Youtube coding tutorial videos", 2017 IEEE 25th International Conference on Program Comprehension (ICPC)

[3] Nabeela Altrabsheh, Mohamed Medhat Gaber, Mihaela Cocea, "SA-E: Sentiment Analysis for Education", The 5th KES International Conference on Intelligent Decision Technologies (KES-IDT), At Portugal, Volume: 255