Automated Sentiment Analysis of Web Multimedia

Ashutosh Bhawsar[1], Devashish Katoriya[2], Ninad Kapadnis[3], Bhushan Shilawat[4], Prof. Anand Kolapkar[5]
[1,2,3,4,5]*Department of Computer Engineering, K. K. Wagh Institute of Engineering Education & Research, Nashik*

***Abstract* - The popularization web-based multimedia content has raised the need to analyze and retrieve it automatically. There is an immense need for identification as well as classification of sentiments (human emotions) due to unfeasibility of labelling big data manually on a larger scale. The Proposed system is to automate the content identification using Audio Processing and Machine Learning. The system aims to extract the audio stream from any multimedia as the input. Using natural language processing, the system will automatically generate raw text useful for identification of the type of multimedia. Further using this data and machine learning, the system will label the content with the opinion of the speaker as output. It will show various sentiments along with their intensity.**

***Index Terms* - sentiment analysis, natural language processing, speech recognition, information retrieval, machine learning, user experience.**

I.INTRODUCTION

Over the last decade, the world has experienced rapid changes. Life has become modern, and the people of the world have to thank the immense contribution of internet technology for communication and information sharing. The internet has emerged as a global encyclopedia of information. Any kind of information on any topic under the sun is available on the internet, including videos and audios.

However, making money from web multimedia leads to a race of content creators to rush the media to get more views. One is thus trapping the user to watch unnecessary and irrelevant content. False advertising is known as click-bait. This leads to a need for verification of media present on the internet.

II. LITERATURE SURVEY

According to the Authors, the system implements a new architecture for audio classification using hybrid keyword spotting. Additionally, two databases were used for audio-based sentiment detection, namely, YouTube sentiment database (7.5hrs, 66 Videos) and another called UT-Opinion Opinion audio archive [1]. The result of ViTS(Video Tagging System) was that each video was indexed with a rich set of labels and linked with other related contents [2].

Authors developed a system that aims to identify aspects on which users' comments. The same system seeks to remove the unnecessary information for analysis, thus shortening a big and complex sentiment sentence into one that is easier to parse [4].

The authors used Machine learning methods to construct emotional arcs in movies to calculate families of arcs and then demonstrated the ability to predict audience engagement. Their results showed that the emotional arcs learned by the approach successfully represent macroscopic aspects of a video story that drive audience engagement. Future work suggested by them was to implement machine understanding to predict audience reactions to video stories and hence improving our ability to better communicate with each other [5].

III.PROPOSED SYSTEM

The proposed system aims to classify the sentiments for web multimedia specifically containing the audio as whole or sub-part. This mainly includes videos and podcasts. Total 8 sentiments are taken into consideration as final output or classes which covers almost every human emotion. List of those classes is as follows - Joy, Sadness, Love, Anger, Trust, Disgust, Surprise, Fear. The input to our system is a web URL that is fetched through a web browser extension. The output contains the three most probable sentiments out of 8 along with their intensity and probability.

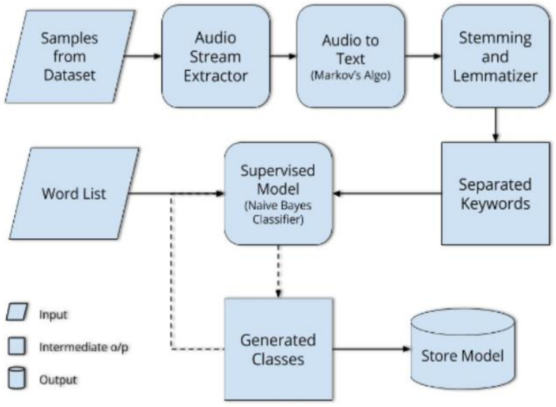


Fig. 1. Working model of Proposed System

IV.TOOLS AND TECHNOLOGIES

A. *Naive Bayes Classifier*

Naive Bayes is a family of probabilistic algorithms that take advantage of probability theory and Bayes' theorem to predict the tag of a text.

For example, "A great game" - Sports, "It was a close election" - Not sports.

It is probabilistic, meaning it needs to estimate the number of occurrences of single event w.r.t the total possible occurrences for each tag in an input text, and thus giving the output of tag having the highest chance. This activity or calculation is achieved with the use of Bayes' Theorem, which describes the probability of a feature, based on the knowledge of features affecting or associated with a particular feature beforehand.

The first thing 'Naive Bayes' does is decide what to use as features. It uses Word Frequencies for this purpose. The advantage is that it does not use sentence ordering, treating every sentence as its own. Stop words need to be removed, and also terms are to be stemmed to root form. Probability in Naive Bayes classifier is calculated as,

P(A|B) = $\frac{P(B|A) \times P(A)}{P(B)}$

P(A|B) = Dependent probability calculated for event A on event B.

P(B|A) = Dependent probability calculated for event B on event A.

P(A) = Independent probability for event A.

P(B) = Independent probability for event B.

B. *Hidden Markov Model*

Hidden Markov Model is used to convert audio speech into text. Four main steps are carried out to convert input speech to text output. These steps are speech database, pre-processing, feature extraction and text recognition. Firstly, the speech database is loaded. The speech signals at low frequencies have more energy than at high frequencies. Therefore, the strength of the signal is necessary to be boosted at high frequencies. The environment's saturation is directly proportional to rate of media recognition, which in turn worsens the accuracy due to unwanted noise. Thus, noise suppression must be performed. The state of pre-processing is followed by the process of extraction of the speech samples, also called as coefficients, using Mel Frequency Cepstral Coefficient (MFCC). Finally, these MFCC coefficients are used as the input of the Hidden Markov Model (HMM) recognizer to classify the desired spoken word.

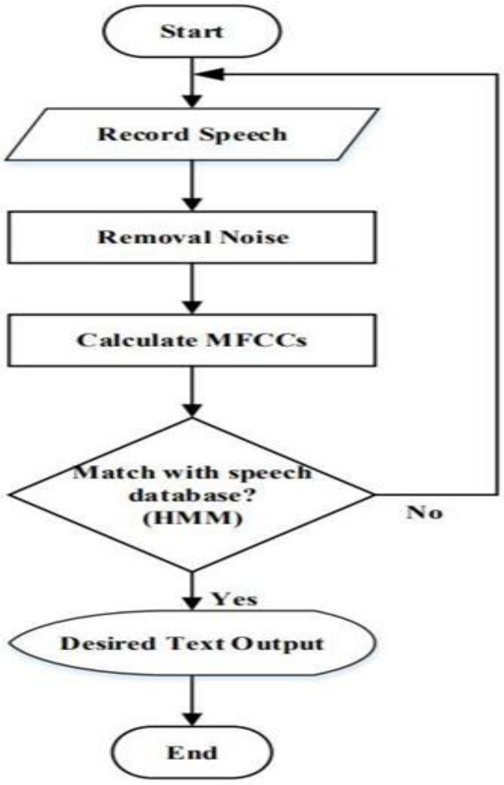


Fig.2-Flowchart of speech to text conversion

Mathematical model is given as,

E = $\sum_{x=1}^{N} \quad log(s(x) \quad ^2)$

ZCR = $\frac{1}{2}\sum_{x=1}^{N} \quad |sgn[s(x+1)] - sgn[s(x)]|$

sgn[s(x)] = +1, if s(x)

sgn[s(x)] = -1, if s(x) < 0

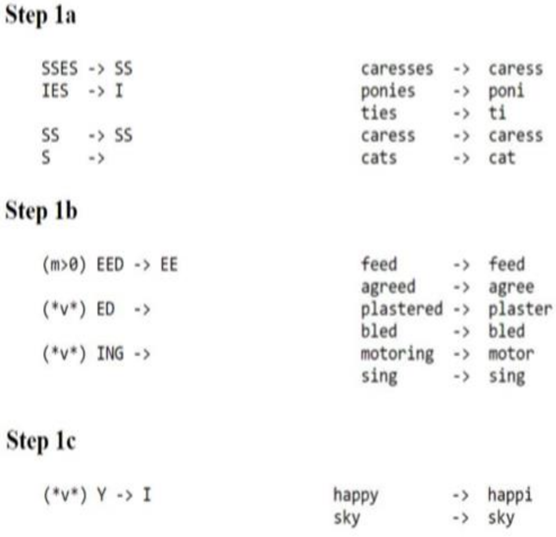
where,

x = speech signal

E = logarithmic short-term energy
ZCR = short-term zero crossing rate

C. Porter Stemmer Algorithm

A word is a combination of two things: A Stem and Affix(es). Stemming is the process of determining the stem of a given work.
For example, generalizations = general + ization + s.
Porter's Stemmer is a rule-based algorithm and has a practical approach. Rules are of the form (condition) S1 -> S2, where S1 and S2 are Affixes and state on S1 is that longest matching affix is taken into consideration. The rules are divided into sets, and in each successive step, a group of controls is applied. Porter Stemmer rules are as follows:



Step 1a

```
SSES -> SS                    caresses  -> caress
IES  -> I                     ponies    -> poni
                              ties      -> ti
SS   -> SS                    caress    -> caress
S    ->                       cats      -> cat
```

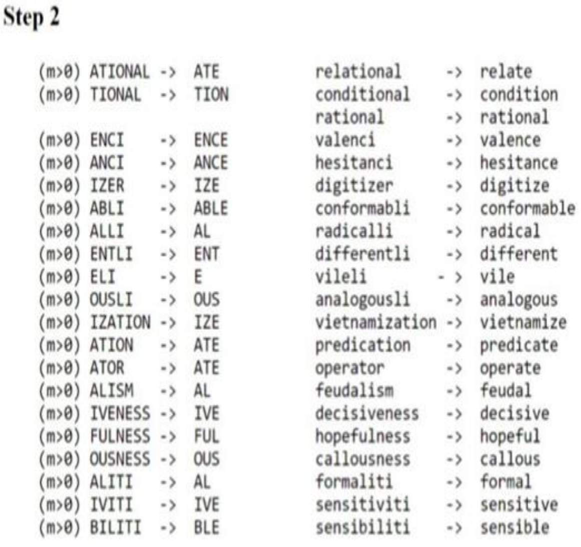
Step 1b

```
(m>0) EED -> EE               feed      -> feed
                              agreed    -> agree
(*v*) ED  ->                  plastered -> plaster
                              bled      -> bled
(*v*) ING ->                  motoring  -> motor
                              sing      -> sing
```

Step 1c

```
(*v*) Y -> I                happy       -> happi
                            sky         -> sky
```

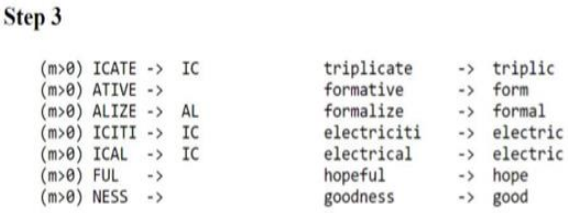
Fig. 3. Step 1 of rules for Porter Stemmer



Step 2

```
(m>0) ATIONAL ->  ATE      relational     ->  relate
(m>0) TIONAL  ->  TION     conditional    ->  condition
                           rational       ->  rational
(m>0) ENCI    ->  ENCE     valenci        ->  valence
(m>0) ANCI    ->  ANCE     hesitanci      ->  hesitance
(m>0) IZER    ->  IZE      digitizer      ->  digitize
(m>0) ABLI    ->  ABLE     conformabli    ->  conformable
(m>0) ALLI    ->  AL       radicalli      ->  radical
(m>0) ENTLI   ->  ENT      differentli    ->  different
(m>0) ELI     ->  E        vileli         - > vile
(m>0) OUSLI   ->  OUS      analogousli    ->  analogous
(m>0) IZATION ->  IZE      vietnamization ->  vietnamize
(m>0) ATION   ->  ATE      predication    ->  predicate
(m>0) ATOR    ->  ATE      operator       ->  operate
(m>0) ALISM   ->  AL       feudalism      ->  feudal
(m>0) IVENESS ->  IVE      decisiveness   ->  decisive
(m>0) FULNESS ->  FUL      hopefulness    ->  hopeful
(m>0) OUSNESS ->  OUS      callousness    ->  callous
(m>0) ALITI   ->  AL       formaliti      ->  formal
(m>0) IVITI   ->  IVE      sensitiviti    ->  sensitive
(m>0) BILITI  ->  BLE      sensibiliti    ->  sensible
```

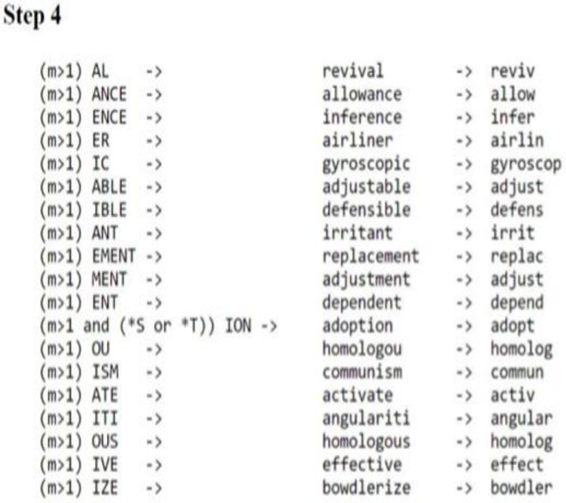
Fig. 4. Step 2 of rules for Porter Stemmer



Step 3

```
(m>0) ICATE ->  IC          triplicate     ->  triplic
(m>0) ATIVE ->              formative      ->  form
(m>0) ALIZE ->  AL          formalize      ->  formal
(m>0) ICITI ->  IC          electriciti    ->  electric
(m>0) ICAL  ->  IC          electrical     ->  electric
(m>0) FUL   ->              hopeful        ->  hope
(m>0) NESS  ->              goodness       ->  good
```

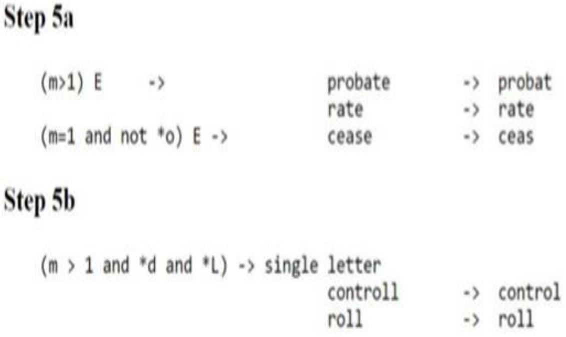
Fig. 5. Step 3 of rules for Porter Stemmer



Step 4

```
(m>1) AL    ->              revival        ->  reviv
(m>1) ANCE  ->              allowance      ->  allow
(m>1) ENCE  ->              inference      ->  infer
(m>1) ER    ->              airliner       ->  airlin
(m>1) IC    ->              gyroscopic     ->  gyroscop
(m>1) ABLE  ->              adjustable     ->  adjust
(m>1) IBLE  ->              defensible     ->  defens
(m>1) ANT   ->              irritant       ->  irrit
(m>1) EMENT ->              replacement    ->  replac
(m>1) MENT  ->              adjustment     ->  adjust
(m>1) ENT   ->              dependent      ->  depend
(m>1 and (*S or *T)) ION -> adoption       ->  adopt
(m>1) OU    ->              homologou      ->  homolog
(m>1) ISM   ->              communism      ->  commun
(m>1) ATE   ->              activate       ->  activ
(m>1) ITI   ->              angulariti     ->  angular
(m>1) OUS   ->              homologous     ->  homolog
(m>1) IVE   ->              effective      ->  effect
(m>1) IZE   ->              bowdlerize     ->  bowdler
```

Fig. 6. Step 4 of rules for Porter Stemmer



Step 5a

```
(m>1) E     ->              probate        ->  probat
                            rate           ->  rate
(m=1 and not *o) E ->       cease          ->  ceas
```

Step 5b

```
(m > 1 and *d and *L) -> single letter
                            controll       ->  control
                            roll           ->  roll
```

Fig. 7. Step 5 of rules for Porter Stemmer

Generally, rules v/s time trade-off is considered. More rules give more accuracy but is slow and vice versa. This slow response is especially crucial in real-time applications like Web Search Engines.

V. RESULTS

We tested our system for accuracy by varying the duration of audio processed for the same set of videos and podcasts combined. Following graph shows the results where X-axis denotes ‘Time duration (in seconds)’ and Y-axis represents ‘Accuracy (in percentage)’.

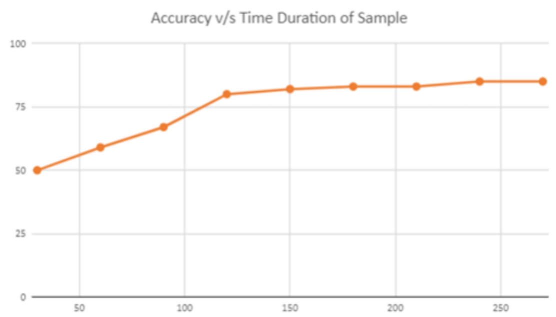


Fig. 8. Result 1 – Accuracy v/s Time duration

Also, we tested our system for the time taken to classify the sentiments in real-time application. Following graph shows the results where X-axis denotes ‘Time duration (in seconds)’ and Y-axis represents ‘Time required (in seconds)’.

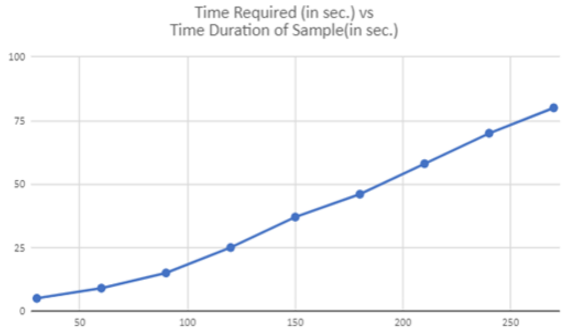


Fig. 9. Result 2 – Time required v/s Time duration

VI. CONCLUSION

The existing system produces search results based on title matching. This way of searching restricts the user to find the appropriate content required as per his needs flexibly. The proposed system identifies sentiments based on the content of the media. Thus, the user gets a general idea about web multimedia before watching it. Hence, the system delivers eight different classes of human emotions, using machine learning and natural language processing. System fetched queries and media, and appropriate emotions get generated.

VII. FUTURE SCOPE

Future work includes classifying the web multimedia into more than eight above mentioned human emotion categories. This addition will increase the generated classes and better allow the user to identify particular media. Also, multiple classifiers like Support Vector Machines and Naïve Bayes can be used together to yield more accurate results. Multiple classifiers can be used together in an ensemble.

VIII. ACKNOWLEDGMENT

We are very thankful to our guide Prof. Anand V. Kolapkar, KKWIEER, Nashik for the guidance we needed and his indispensable support, suggestions.

REFERENCES

[1] Lakshmish Kaushik, Abhijeet Sangwan, and John H. L. Hansen, "Automatic Sentiment Detection in Naturalistic Audio", IEEE TRANSACTIONS ON AUDIO AND LANGUAGE PROCESSING, VOL. 25, NO. 8, AUGUST 2017.

[2] Delia Fernandez, David Varas, Joan Espadaler, Issei Masuda, Jordi Ferreira, Alejandro Woodward, David Rodriguez, Xavier Giro Nieto, Juan Carlos Rivero and Elisenda Bau, "Video Tagging System from Massive Web Multimedia Collections (ViTS)", 2017 IEEE International Conference on Computer Vision Workshops.

[3] Harika Abburi, Manish Shrivastava and Suryakanth V Gangashetty, “Improved Multimodal Sentiment Detection Using Stressed Regions of Audio", 2016 IEEE Region 10 Conference (TENCON) — Proceedings of the International Conference.

[4] Wanxiang Che, Yanya Zhao, Hongli Guo, Zhong Su, and Ting Leu, "Sentence Compression for Aspect-Based Sentiment Analysis", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 12, DECEMBER 2015.

[5] Eric Chu, Deb Roy, "Audio-Visual Sentiment Analysis for Learning Emotional Arcs in Movies", 2017 IEEE International Conference on Data Mining.

[6] Julio Savigny, Aayu Purwareanti, School of Electrical and Informatic Engineering, Institute Teknologi, Bandung, Indonesia, "Emotion Classification on You Tube Comments using Word Embedding", 978-1-5386-3001-3/17.

[7] Preity, Sunny Dahiya, "SENTIMENT ANALYSIS USING S.V.M. AND NAIVE BAYES ALGORITHM", IJCSMC, Vol. 4, Issue. 9, September 2015, pg.212 – 219.