SENTIMENT ANALYSIS ON TRENDING YOUTUBE VIDEO BY USER COMMENT

[1]Anamika, [2]Dr.Megha Kamble

Research Scholar, Department of Computer Science and Engineering Lakshmi Narain
College of Technology Bhopal (M.P)
Assistant Professor, Department of Computer Science and Engineering Lakshmi Narain
College of Technology Bhopal (M.P)

Abstract

User comments are the most popular but also extremely controversial form of communication on YouTube. Opinion mining or comment toward attitude evaluation, individual entity, are usually called sentiment. Everyone is free to give opinion related with the present opinions on youtube. Hence people have a free will to express their opinion regarding the performance. Due to the raise of many critics that appear in a short amount of time, there a needs to conduct research on opinion mining. Sentiment analysis is a technique used by researchers to measure and classify the popular content from social media. Public sentiments related to prediction of events, such as Election demonstrations, indicate public attitude, public interest and predict the election results. In this research, opinion mining is applied on the Youtube comments related to "PRIME MINISTER NARENDRA MODI'S MANN KI BAAT WITH THE NATION, AUGUST 2020". using machine learning techniques. The proposed model will show a good ability to predict user sentiments.

Keywords: Sentiment analysis, youtube, Opinion mining,SVM.

1.1   Introduction

In Social media channels, such as Twitter, Facebook, LinkedIn, Instagram and Google+ allow for people to share their views and opinions about any public topic. Social media monitoring has been growing day by day for analyzing and classifying of social media data. It plays an important role for people's opinions. Sentiment analysis or opinion mining is a technique used by several researchers to measure and classify the popular content from social media. In simple terms, opinion mining involves making a system to use of reviews posted by user so as to improve upon the products features. Given a set of reviews our task involves the following: Identify features of the product on which customers have expressed their opinion (called product features). We use techniques like data mining and natural language processing in order to mine the features. [1]

For each feature, partition the entire set of reviews into positive or negative reviews. To decide what the opinion orientation is we perform the three ssubtasks:

1. Identify a set of adjectives normally used to express opinions using natural language processing method. These are called opinion words.
2. For each opinion word we determine its semantic orientation.

Decide the opinion orientation for each sentence generating a summary out of the discovered information.

The need of Sentiment Analysis or opinion mining refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Opinion mining or Sentiment Analysis involves analyzing opinions, sentiments or mentality of the writer from the written text. Opinion mining uses the concepts of data mining and machine learning to perform this task. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication. Everyday enormous amount of data is created from social networks, blogs and other media and diffused in to the World Wide Web [2]

1.2 Applications of Sentiment Analysis In Real World

A. Product and service reviews:
Reviews of consumer products and service. Many automated websites provides feedback. Eg: Google product search.

B. Reputation Monitoring:
Monitoring reputation of a specific brand.Eg: Twitter, Facebook.

C. Election Prediction:
By analyzing sentiments from different sources one can predict the outcome of the event. Eg: Election. It enables managers to track how voters feel about different issues, how they relate to speeches and actions of the candidate.

D. Decision making:
Sentiment analysis uses these various sources to find the articles that discuss the aggregate score. Eg. The Peoples opinion on GST graphically shows positive and negative sentiment of each Sentence.

1.3 Problem of Statement

Predicting the popularity of content is a complex and difficult task and different prediction methods and strategies have been proposed in several recent studies. Sentiment analysis relates to the problem of mining the sentiments from online available unstructured data and categorizing the opinion expressed by a people towards particular events into at most three preset categories: positive, negative and neutral and these classification methods plays very crucial role for unstructured data management. There is a need to classify the twitter data using sentiment analysis techniques.

1.4 Literature Review

. Cagatay Catal & Mehmet Nangir (2017) analyzed the potential advantages of multiple classifier system and presented Sentiment Classification Model. Vote algorithm was employed in combination with three classifiers namely, Naive Bayes, Support Vector Machine (SVM), and Bagging was compared with multiple classifier systems. Multiple

classifier systems helped in enhancing the performance of individual classifiers and Meta classifiers. However, methods for increasing the true positive rate were not considered.

Danushka Bollegala et al. (2013) developed cross-domain sentiment classifier using an automatically extracted sentiment sensitive thesaurus. For both source and target domains to resolve the issues of crossdomain sentiment classification, a sentiment sensitive distributional thesaurus was implemented with the aid of labeled data and unlabeled data. Sentiment sensitivity was improved by the integration of document level sentiment labels in context vectors utilized as the origin for evaluating the distributional similarity among words. Then, thesaurus was generated for increasing feature vectors in binary classifier. Cross-domain sentiment classifier performed better when compared with other conventional cross-domain sentiment classification methods. But, cross-domain sentiment classifier failed to improve the classification accuracy.

Anindya Ghose & Panagiotis G Ipeirotis (2011) analyzed the effects on economic results namely product sales and observed the consequences of social results. A multiple characteristic of review text namely, subjectivity levels, different measures of readability and coverage of spelling errors for predicting significant text-based features was investigated.

Oscar Araque et al. (2017) described a deep learning based sentiment classifier with the aid of word embedding model and linear machine learning algorithm. For sentiment analysis, two ensemble methods were designed to aggregate the baseline classifier with another surface classifiers and an integration of both surface and deep features for joining information from different sources. Then, taxonomy was introduced to categorize the models to recognize the performance. But, deep learning based sentiment classifier failed in addressing the issues of error rate relative to truncation.

Kotelnikov & Pletneva (2016) described text sentiment analysis method depending on the weights of sentiment words. Text sentiment analysis method was employed for identifying negative or positive sentiment according to certain object. Then by using co-clustering in Text sentiment analysis method, search space of optimal weights of sentiment words was solved. Co-clustering helped in attaining related sentiment words group and text documents. Depending on genetic algorithm for each cluster, the weights were optimized. Text sentiment analysis method was efficient in attaining compact lexicons for explaining the collection of text document. However, classification accuracy was reduced in text sentiment analysis method.

1.5 Implementation

The goal of this section is acquisition of comment of a selected YouTube video from "PRIME MINISTER NARENDRA MODI'S MANN KI BAAT WITH THE NATION, AUGUST 2020". In order to address this task a focused crawler is implemented. According to the video URL, it extracts comment (up to 1000) of that video using web API through HTTP GET method. But, the extracted comments are heterogeneous in terms of languages

and various notions used by the users. Therefore, we carried out some preprocessing on these unstructured comments to generate the data sets.

After extracting the comments, following changes are performed:

- Remove all the expressions which are irrelevant for the proposed methodology like date ("Dec 2- 2010"or "2-12-2010"), link (www.imdb.com, www.tmdb.com etc.), numbers (12, 20 etc.) and special characters ("*","/","!","@","?","#","&","$"), emoticons ("☺, ☹,<3 etc.) and different language (Chinese, Arabic, Bangla , Hindi etc.).
- Remove all the punctuations such as period ("."), space ("-"), commas (","), semicolon (";"), hash ("-") etc.

Table1 Retrieved Data detail

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Comment	AuthorDis	AuthorPro	AuthorCh	AuthorCh	ReplyCou	LikeCount	Published	UpdatedA	Comment	ParentID	VideoID	
2	Amit Shah	AJ Ahuja	https://yt	http://ww	UCJclepx4	0	0	2020-10-0	2020-10-0	UgyUxNCe	NA	OExRBGfhLKc	
3	Man ki bat	Simran Ka	https://yt	http://ww	UCSJ1r0dr	0	1	2020-10-0	2020-10-0	UgzGt8nel	NA	OExRBGfhLKc	
4	Bank railw	Simran Ka	https://yt	http://ww	UCSJ1r0dr	0	1	2020-10-0	2020-10-0	UgxoROd>	NA	OExRBGfhLKc	
5	On behalf	ICMC HOP	https://yt	http://ww	UCVo85aG	0	0	2020-10-0	2020-10-0	Ugzwy6Pz	NA	OExRBGfhLKc	
6	Dislike fu	J O T C H A	https://yt	http://ww	UCCHtnOc	0	0	2020-10-0	2020-10-0	Ugzyq6hh	NA	OExRBGfhLKc	
7	No. 1 wro	Niladri Sh	https://yt	http://ww	UCFvc_AI2	0	1	2020-10-0	2020-10-0	UgxQiKxI(	NA	OExRBGfhLKc	
8	Modi dest	Deepak M	https://yt	http://ww	UCd-U6eC	0	1	2020-10-0	2020-10-0	UgyuwLd8	NA	OExRBGfhLKc	
9	Beta karal	Deepak M	https://yt	http://ww	UCd-U6eC	0	0	2020-10-0	2020-10-0	UgwRqICh	NA	OExRBGfhLKc	
10	2024 ke ba	Sandeep r	https://yt	http://ww	UCBaI1Q7	0	0	2020-10-0	2020-10-0	Ugwwrfb4	NA	OExRBGfhLKc	
11	Are ja tu b	Sandeep r	https://yt	http://ww	UCBaI1Q7	0	0	2020-10-0	2020-10-0	UgyTBLDS	NA	OExRBGfhLKc	
12	Hathrash	BALRAM F	https://yt	http://ww	UC0sThTd	0	0	2020-10-0	2020-10-0	UgxNifr25	NA	OExRBGfhLKc	
13	feku	Swapnil B	https://yt	http://ww	UCcr7DOd	0	1	2020-10-0	2020-10-0	UgyEDKuE	NA	OExRBGfhLKc	
14	ashaamee	Lakhveer !	https://yt	http://ww	UCdFbxon	0	0	2020-10-0	2020-10-0	Ugyf6rbd>	NA	OExRBGfhLKc	
15	<U+0001F4	Lakhveer !	https://yt	http://ww	UCdFbxon	0	0	2020-10-0	2020-10-0	UgyQ9AG(	NA	OExRBGfhLKc	
16	<U+0930>	Abhi Vish	https://yt	http://ww	UCZ3LQKv	0	0	2020-09-3	2020-09-3	UgwmgSy	NA	OExRBGfhLKc	
17	Rahiiyeaa	Mok YouT	https://yt	http://ww	UCFRzWC	0	0	2020-09-3	2020-09-3	UgzIn6Okl	NA	OExRBGfhLKc	
18	Make in b	Ankit Raj	https://yt	http://ww	UCljc5jioy	0	1	2020-09-3	2020-09-3	UgyVHOFI	NA	OExRBGfhLKc	
19	Big dislike	Nilutpal S	https://yt	http://ww	UCHAHifp	0	3	2020-09-3	2020-09-3	UgzWXZv	NA	OExRBGfhLKc	
20	<U+092C>	built kids	https://yt	http://ww	UCFA8YVz	0	1	2020-09-3	2020-09-3	Ugyt9Ls-ft	NA	OExRBGfhLKc	
21	J U S T I C I	Middle Cl	https://yt	http://ww	UC_bEabU	0	1	2020-09-3	2020-09-3	Ugy6-fz00	NA	OExRBGfhLKc	
22	Kasam sa	T.D gamin	https://yt	http://ww	UCR4CI8O	0	1	2020-09-3	2020-09-3	Ugwkq01(	NA	OExRBGfhLKc	
23	Sirf modi	T.D gamin	https://yt	http://ww	UCR4CI8O	0	0	2020-09-3	2020-09-3	Ugy0Xb9D	NA	OExRBGfhLKc	
24	Inquilab Z	Red kite	https://yt	http://ww	UC-CbhfV	0	1	2020-09-3	2020-09-3	Ugxoy7AV	NA	OExRBGfhLKc	
25	Tann ki ba	SRB's univ	https://yt	http://ww	UCp52WD	0	1	2020-09-3	2020-09-3	UgxqN5q(	NA	OExRBGfhLKc	

1.6   Result

Total video comments extracted: 3859

- Getting Youtube comments from youtube Api with the help of api key.
- Deriving the sentiment of each comment: Sentiment for each commentwas computed based on the sentiment score of the terms in the comment. The sentiment of a comment is equivalent to the sum of the sentiment scores for each term in the comment. The _le nrc . containing a list of pre-computed sentiment scores was used to determine term score in each tweet.
- Deriving the sentiment of new terms: The sentiment for the terms that do not appear in the file nrc was computed based on the overall tweet sentiment deduced in step 2.
- Computing term frequency: Frequency of each term was calculated as of occurrences of the term in all comments]/[of occurrences of allterms in all comments].
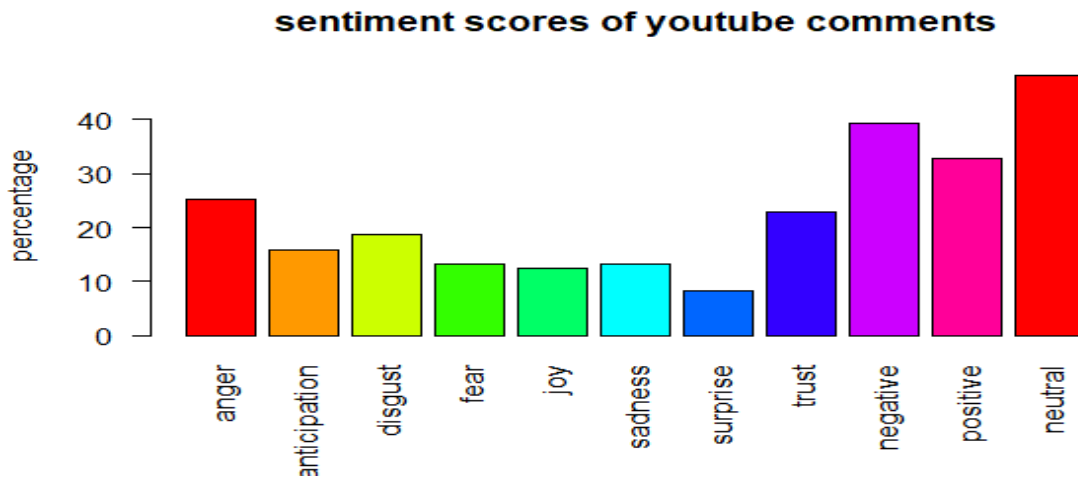
Graph 1 Sentiment score of youtube comments

The calculation of accuracy value of analysis towards the SVM method's result that was done using need to have the accuracy, precision and recall performance evaluation from the experiment with confusion matrix method. The evaluation is done using Confusion Matrix includes the True Positive Rate (TP rate), True Negative Rate (TN Rate), False Positive Rate (FP Rate) and False Negative Rate (FN rate) as an indicator. The TP rate is the percentage of the positive class which was classified as the positive class, whereas the TN rate is the percentage of the class negatively classified as a negative class. FP rate is class negative which is classified as positive class. The FN rate is a class positive that is classified as a negative class. [7]

Table 0.2 Accuracy Performance Measure

Methodology	Accuracy (%)
Existing work [25]	75.435%
Proposed work	86%
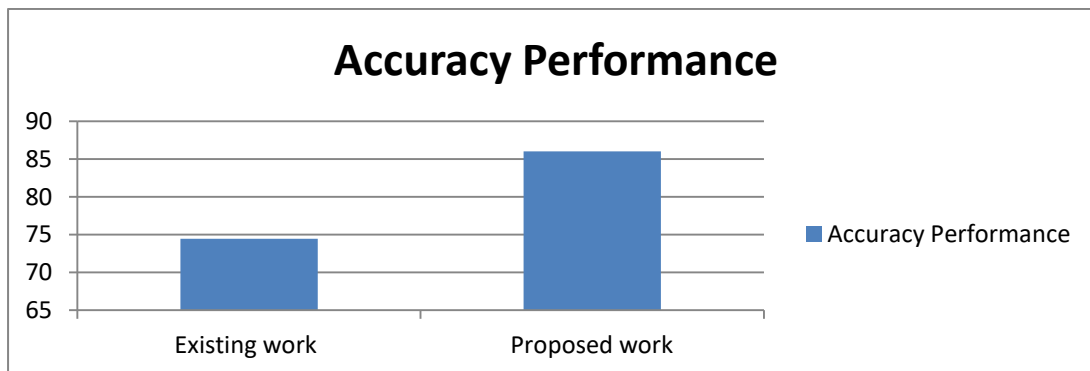
Graph 2Accuracy Performance

1.7 Conclusion

Classification of general events and detection of Sentiment Polarity of user comments in YouTube is a challenging task for researchers so far. A lot of work is done in this regard but still have a long way to go to overcome this problem. In this work we have emphasized on following problems in order to find the polarity of comments given by the users of YOUTUBE.

1) Current sentiment dictionaries having limitations.

2) Informal language styles used by users,

3) Estimation of sentiments for community-created terms,

4) To assign proper labels to events,

5) Achieve satisfactory classification performance

6) Challenges involving social media sentiment analysis.

References

1 Kerstin Denecke and Yihan Deng. Sentiment analysis in medical settings: New opportunities and challenges. Arti_cial Intelligence in Medicine, 2015.

2 Alessia D'Andrea Fernando Ferri," Approaches, Tools and Applications for Sentiment Analysis Implementation", International Journal of Computer Applications (0975 – 8887) Volume 125 – No.3,September 2015.

3 Cagatay Catal & Mehmet Nangir,,A sentiment classification model based on multiple classifiers", Applied Soft Computing, Elsevier, vol. 50, pp. 135–141. 2017

4 Danushka Bollegala, David Weir & John Carroll, Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus", IEEE Transactions on Knowledge and Data Engineering, vol. 25, issue no. 8, pp. 1719-1731. 2013

5 Oscar Araque, Ignacio Corcuera-Platas, J.Fernando Sánchez-Rada & CarlosA. Iglesias, Enhancing deep learning sentiment analysis with ensemble techniques in social applications", Expert Systems with Applications, Elsevier, vol. 77, pp. 236-246. 2017

6 Kotelnikov, EV & Pletneva, MV, ,,Text sentiment classification based on a genetic algorithm and word and document co-clustering", Journal of Computer and Systems Sciences International, vol. 55, issue no. 1, pp. 106–114. 2016.

7 LuhRia Atmarani., I.A. Dwi Giriantari., Made Sudarma., "Sistem Opinion Mining dengan Metode Pos Tagging dan SVM Untuk Ekstraksi Data Opini Publikpada Layanan JKBM" Teknologi Elektro, Vol. 16, No1, 2017.

8 Yang Liu, Jian-Wu Bi & Zhi-Ping Fan,Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, Information Fusion, Elsevier, vol. 36, pp. 149–161. 2016

9 Anna Jurek, Maurice D Mulvenna & Yaxin Bi „Improved lexicon based sentiment analysis for social media analytics‟, Security Informatics, Springer, vol. 4, issue no. 9,pp. 1-13. ,2015.

10 Chuanjun Zhao, Suge Wang & Deyu Li,Determining Fuzzy Membership for Sentiment Classification: A Three-Layer Sentiment Propagation Model‟, Plos One, vol. 11, issue no. 11, pp. 1-32. 2016

11 Mauro Dragoni, Fondazione Bruno Kessler Giulio Petrucci & Fondazione Bruno Kessler, A Neural Word Embeddings Approach For Multi-Domain Sentiment Analysis‟, IEEE Transactions on Affective Computing, vol. 8, issue no. 4, pp. 457 – 470. 2017.

12 Alpaslan Burak Eliacik & Nadia ErdoganInfluential,User Weighted Sentiment Analysis on Topic Based Microblogging Community‟, Expert Systems with Applications, Elsevier, vol. 92, pp. 403-418. 2018.

13 Gilad Katz, Nir Ofek & Bracha Shapira, ConSent: Context-based sentiment analysis‟, Knowledge-Based Systems, Elsevier, vol. 84, pp.162-178. 2015.

14 Vandana Jha, Savitha R, Deepa Shenoy, P, Venugopal, KR & Arun Kumar Sangaiah,A novel sentiment aware dictionary for multidomain sentiment classification‟, Computers & Electrical Engineering, pp. 1-13. 2017

15 Nadia FF da Silva, Eduardo R Hruschka & Estevam R Hruschka Jr, „Tweet sentiment analysis with classifier ensembles‟, Decision Support Systems, Elsevier, vol. 66, pp. 170-179. 2014.