

An Efficient Framework for Detecting Various Moods in Hinglish and English Dataset

Vikas Tripathi, Himanshu Silswal, Gaurav Rawat, Tanmay Jain



Abstract: Due to the paced growth in web technologies and natural language processing, research on Sentiment Analysis (SA) has become very popular in recent times. In recent years most of the research papers have focused on sentiment analysis based on polarity (positive and negative sentiments). This paper presents an effective framework for identification of various moods of person from its written text or sentences. The paper focuses on the mood detection in given text written in mixed language called "Hinglish". Hinglish is actually a fusion of two languages, English with the Hindi language. The major goal of this research is to propose a methodology for extracting information of emotions from a given text in Hinglish. The framework tested on 700 sentences containing Hinglish data. Seven emotions anger, happy, joy, confidence, sadness, tentativeness and fear have been used for generation of results. The proposed approach yielded an accuracy of 93.96%.

Keywords: Hinglish, Mixed Language, Natural Language Processing, Sentiment Analysis.

I. INTRODUCTION

The process of classifying the opinions, that are expressed in a piece of text technically or computationally, especially to determine whether the writer's attitude is positive, negative, or neutral towards a particular topic, product, is called sentiment analysis (SA). It is a natural language processing (NLP) task that is used to extract opinions or emotion or sentiments from the text for a given topic. [1][2] Today social media is not only used for conversing, socializing and file sharing but has gone much beyond that. Due to the rapid evolvement of social media, everyone is connected through various social media platforms. S.A recognize the expression or phrase in a text that contains some moods. Based on people's sentiments, potential livable places are analyzed to find accuracy. Of the works discussed, the analysis of the sentiments of a customer for a product through reviews has been the most approached area to analyze sentiments due to its potential in both increasing accuracy and business benefits.

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Vikas Tripathi*, Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India. Email: vikastripathi.be@gmail.com

Himanshu Silswal, Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India. Email: himanshusilswal5@gmail.com

Gaurav Rawat, Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India. Email: RawatGaurav360@gmail.com

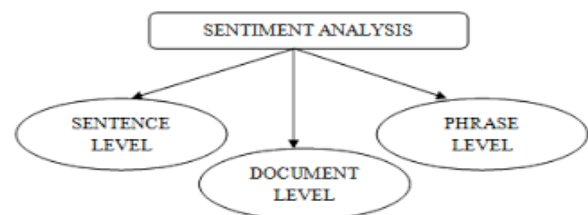
Tanmay Jain, Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India. Email: csetanmayjain@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

The author may speak about subjective opinions/objective facts. It is mandatory to differentiate between these two.

S.A finds the subject towards whom the sentiment is directed. It is necessary to find entity towards which the emotion is directed

from the text that may contain many entities. Many organizations use social media as a tool to understand the likes and dislikes of their customers. This can be done through sentiment analysis or opinion mining as well. Sentiment analysis involves many tasks such as subject-related detection, text preprocessing, and feature extraction and emotion classification. [2] Analyzing sentiments in text data sources like emails or social media posts provides businesses with insights to learn about customers' decisions and behavior. Besides identifying the opinion, these systems



extract attributes of the expression too: Positive or negative opinion expressed by the speaker the thing that is being talked about, the person that expresses the opinion. There are three

Figure.1. Different level of sentiments

different levels of sentiment analysis.

Figure.1. above shows the different phases of S.A. Document-level is the one in which sentiments are drawn out from the reviews, and a whole emotion or thought process is classified, is based on the overall sentiment of the opinion holder. The goal is to classify a review as positive, negative, or neutral. The techniques used for document-level sentiment classification are mostly based on supervised learning. Sentence-level classification considers each sentence as a separate unit and assumes that sentence should contain only one opinion. Sentence-level sentiment analysis has two tasks; subjectivity classification and sentiment classification. Phrase-level classification is the one in which phrases or aspects of a sentence are analyzed to classify as positive or negative. Most of the existing work has been done in the English language but nowadays, in India, there is a lot of increase in 'Hinglish' text. A Hinglish sentence is a combination of Hindi and English languages word. It's a Hindi sentence written using the English language. For example: "ajar ka din baht ache the." The sentence has a literal means in the Hindi language but it is written in English instead of Hindi scripts. It is very important to analyze this type of text as most of the Indians comment and text in this format only.

An Efficient Framework for Detecting Various Moods in Hinglish and English Dataset

A mix-language has been used for analysis herein. In Hindi too, a little amount of work has been done. Hindi is a resource-scarce language. This is also known as mix code emotions analysis.

Now a days, the work on this subject has now became so far, each and every industry, there is the need of the analysis in different-different field. There is the need to develop these types of framework that is needed to do this work of analysing the data. We have preferred to do sentiment analysis of the given datasets.

II. LITRATURE REVIEW

For the proper taxonomy of emotions or sentiments, during past few years there were many researchers that have made it possible to create and apply different deep learning and machine learning technique to achieve the desired result of sentiment analysis. This part will describe in brief, it has the focus on various research related to mood detection analysis, that are used to judge or detect the emotions as well as, contents about different user view point, evaluation in the direction of different subjects and result of using different ML techniques. Emotion analysis or sentiments is an enterprise that can be done adequately by applying different models such as deep learning models, which have been expanded recently. These prototypes or models incorporate, deep learning (DL) [3][4][5][6] method are used and convolutional neural network (CNN) [7][8][9] recursive neural network (RNN) [10]. A part from CNN and RNN, many researchers have used more than one model that is the hybrid neural network section. The proposed approach is on sentiment analysis of the Hinglish datasets with our module and sentiments analysis of the English datasets using the IBM Watson tone analyser [11]. Our approach gives us a better accuracy than machine learning technique, by which we can solve the problem for the sentiment analysis of Hinglish dataset. We have got the knowledge from the previous papers, that are on the visual sentiment analysis, in which the visual expression on the given image for mood detection. And some sentiment analysis based on English text, which delivers the polarity like positive, negative and neutral. So that we have proposed the framework that further gives a detailed mood of emotion. Nowadays due to the enhancement of social media and technology, it is mandatory to create a framework that must study the reviews on the different products and to analysis the tweets on the different ideologies of the politician and great personalities, or the product-based company and firms. It is easy to describe the sentiments on English text, but it is difficult to analyse the mood on the Hinglish text. So, we have created the framework that will allow you to analyse the mood on the English as well as Hinglish.

III. METHODOLOGY

The methods which we use in this approach are that we use the concept of data collection, text processing, negation handling etc. The following is our workflow that we have showcased in our flow chart that how our model of mood detection framework and the methodology used: as shown in 'fig. 2' and table 1, the flow of this framework works in a manner that it collects the data in the form of the sentences and split them word wise. There is a method for checking the

data too whether it is in English or Hinglish. If it's in English then it goes to the IBM Watson tone analyzer, else it goes to the mood analyzer (Hinglish model). Further, the framework searches for the main word in the sentence and checks it with our dictionary to predict the mood of the sentences. In the text pre-processing, it split out the entire sentence from data collection to an array. Then a loop will run from zero to the length of the array. In the loop, it will first detect the language of the sentence. If the sentence will be in English, it will go to IBM Watson Ton Analyzer. Else it will go to the Hinglish model. Both will return a mood that will be stored in an array. After the end of the loop, we will make a confusion matrix of all the moods we get and find out the accuracy, recall, and precession. Now when we get the sentence, we will split out the words from it to an array. Then a loop will run from zero to the length of the array. Post that, we will check whether the word is in the dictionary or not. If the word matches, we return the type of mood that belongs to the sentence. As shown in fig. 3, handling of word having negation is performed. If the sentence in the dataset is found to have words like 'not', 'nahi' etc. then the mood of the sentence is changed. For example, for the English sentence, "Today I am not happy ", it will show the mood as sadness. For a Hinglish sentence like this, "mai tujhse nahi darta hu", it will show the mood as confident.

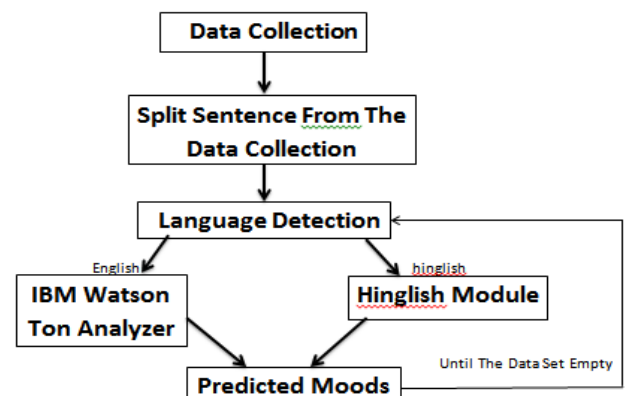


Figure.2. flow of the frame work

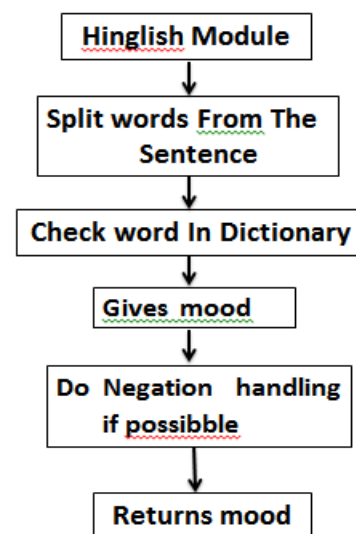


Figure.3. flow of Hinglish module

Table I Algorithm for Working framework

Pseudo code	Abbreviation
Start	Where
Input x	x= sentences.
if x==1, ta(x)	ta=tone analyzer
else if x==0, ha(x)	ha=Hinglish mood analyzer
else print invalid input	Xm=mood predicted
print Xm	
Stop	

IV. RESULT AND DISCUSSION

We use Windows 10 configuration; 64bit, Python Spyder 2.7. In this paper, the framework that we have proposed does the work by taking input - processing it- and then giving output. The sentences are taken as input and check whether the input sentences is Hinglish or English, if it is English then it will send that string to IBM Watson tone analyzer. Else the data goes to the mood detection analyzer. The proposed approach also increases the Hinglish capacity of IBM Watson. We use Windows 10 configuration; 64bit, Python Spyder 2.7. In this paper, the framework that we have proposed does the work by taking input - processing it- and then giving output. The sentences are taken as input and check whether the input sentences is Hinglish or English, if it is English then it will send that string to IBM Watson tone analyzer. Else the data goes to the mood detection analyzer. The proposed approach also increases the Hinglish capacity of IBM Watson.

Table II. accuracy percentage of predicted mood

Accuracy score by Mood	Percentage
Accuracy score by Anger Mood	90
Accuracy score by Confident Mood	99
Accuracy score by Happy Mood	99
Accuracy score by Fear Mood	98
Accuracy score by Joy Mood	84
Accuracy score by Sadness Mood	94
Accuracy score by Tentative Mood	92

Here taken, are the seven different types of moods that are further concepts than the negative, positive and neutral sentiments. Our concepts describe seven different emotions after analyzing the text written in Hinglish. The module also provides the information of the number of the positive, negative and neutral sentences. In the proposed framework, we have used the dictionary mapping concept that maps the text and predicts the sentiments.

Table III. percentage of the framework

Functioning of the framework	Percentage
Accuracy of the framework	93.74%
Precision of the framework	99.03%
Recall of the framework	84.23%

In table II and table III, they represent the accuracy of these different moods and also the accuracy, precision and recall of the framework too. The technique that is used gives us a better accuracy of 93.74%.

table IV. shows the actual mood of the given sentences and the mood predicted by our module for Hinglish datasets. The program has an accuracy of 93.71428571428571 %, that mean it give the exact prediction for the given sentences. No research is without limitations. We had large pool of resources while conducting the research work. It was found that few sentences from the dataset are available for “benchmarking” the algorithm in this context. It has some limitations like it can’t judge the multiple but equal number of the sentiments/moods.

Table IV. working of the framework on Hinglish

Sentences	Actual mood	Predicted mood
Aaj meri tumse ladaai hone ke kaaran main tumse naraj hoon.	Anger	Anger
Hum sabhi ko khush rehna chaiye	Happy	Happy
Tum udas kyun ho, tumhara tumhari girlfriend se jhagda hua hai kya.	Sadness	Sadness
Yaha bahut hi sanjeeda mamla hai.	Tentative	Tentative
Tumhara Jeevan aanandmay ho	Joy	Joy

Table V. Limitation of the framework

Sentences	actual	Predicted	reason
Subaha se vaha gusse mein tha lekin ghumme jaane ki baat se vaha khush ho gya.	happy	Anger	It is because the first mood in the string is anger
Tum sab udas ho, sabhi ko khush rehna chahiye	happy	Sadness	It is because the first mood in the sentences is sadness

In table V, two sentences create a contradiction about the actual state of mood and the predicted state of mood. Thus, the module has limitation that the mood which comes first in the given sentences is the predicted mood which is not as equal as actual mood of the sentences. ”so, we have kept it for the future reference.

V. CONCLUSION AND FUTURE SCOPE

In this paper, the proposed framework has a Hinglish significance and can also detect the mood on the Hinglish text. In the given paper, we have proposed an approach for sentiment analysis of word or sentences in Hinglish text by using dictionary-based technique. Hinglish tweets review dataset has been collected from various sources. These are the many-many applications on mood analysis in different social networking websites, response to a survey, competitions, it is also practical for use in analysis of data for business and different type of situations in which the text is needed to be analyzed.



It can also be used to find whether the people are happy with our products as well as services and their feedback and their experience can promote our products. Computation of a satisfaction metrics for the customer, we can get an idea, how happy are the customers with your products from the ratio of different types of tweets depend on their mood about the product and services. By spotting dissatisfaction or problems with products, it can be used for customer service identifying detractors and promoters. We can do lot of work, regarding the concept of mood detection analysis.



Tanmay Jain perusing B. Tech Computer Science and Engineering from Graphic Era University Keen Learner, Innovation Catalyst, Curious, Interested in R&D in the field of Artificial Intelligence, Data Science Engineering, Machine Learning. Wants to work on innovative ideas, products with a great team that have the potential to impact the lives of millions. Ability to

implement almost anything.

REFERENCES

1. J. Islam and Y. Zhang, Visual Sentiment Analysis for Social Images Using Transfer Learning Approach, 2016 IEEE Int. Conf. Big Data Cloud Comput. (BDCloud), Soc. Comput. Netw. (SocialCom), Sustain. Comput. Commun., pp. 124130, 2016.
2. Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirb, Machine Learning-Based Sentiment Analysis for Twitter Accounts Math. Comput. Appl. 2018, 23, 11
3. P. Vateekul and T. Koomsubha, A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data, 2016.
4. R. Ghosh, K. Ravi, and V. Ravi, A novel deep learning architecture for sentiment classification, 3rd IEEE Int. Conf. Recent Adv. Inf. Technol., pp. 511516, 2016.
5. P. Vateekul and T. Koomsubha, A Study of Sentiment Analysis Using Deep Learning Techniques on Thai Twitter Data, 2016.
6. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, Deep learning for visual understanding: A review, Neurocomputing, vol. 187, pp. 2748, 2016.
7. X. Ouyang, P. Zhou, C. H. Li, and L. Liu, Sentiment Analysis Using Convolutional Neural Network, Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput. (CIT/IUCC/DASC/PICOM), 2015 IEEE Int. Conf., pp. 23592364, 2015.
8. N. Kalchbrenner, E. Grefenstette, and P. Blunsom, A Convolutional Neural Network for Modelling Sentences, Acl, pp. 655665, 2014.
9. Y. Kim, Convolutional Neural Networks for Sentence Classification, Proc. 2014 Conf. Empir. Methods Nat. Lang. Process. (EMNLP 2014), pp. 17461751, 2014.
10. R. Socher and C. Lin, Parsing natural scenes and natural language with recursive neural networks, Proc. , pp. 129136, 2011.
11. <https://github.com/watson-developer-cloud>

AUTHORS PROFILE



Himanshu Silswal perusing B. Tech from Graphic Era University. He is currently active in some research work such as the sentiment analysis using machine learning algorithm, sentiment analysis of Hinglish and English data for more classified form. He is since one year being with this concept of sentiment analysis.



Gaurav Rawat perusing B. Tech from Graphic Era University. He is currently active in some research work such as the sentiment analysis using machine learning algorithm, sentiment analysis of Hinglish and English data for more classified form. He is since one year being with this concept of sentiment analysis.



Dr. Vikas Tripathi has done BE in information technology from Technocrats institute of technology, Bhopal, M. Tech in Software engineering from Indian institute of information technology Gwalior and PhD from Uttarakhand technical university, Dehradun. He is actively involved in research related to Software engineering, Computer Vision, Machine learning and Video Analytics. He has published many papers in reputed international conferences and journals. Currently he is working as an associate professor in Graphic era deemed to be university Dehradun, India..