A Recommendation System For Status Suggestion Using Sentiment Analysis With Social Media

[1]Vishakha N. Gandha, [2]Prof. N. J. Rola

[1]Research Scholar, Department of Computer Engineering, Darshan Institute of Engineering & Technology Rajkot, Gujarat, INDIA
[2]Assistant Professor, Department of Computer Engineering, Darshan Institute of Engineering & Technology Rajkot, Gujarat, INDIA
[1]vishgandha@gmail.com, [2]nitin.rola@darshan.ac.in
_____

Abstract

Now a day's sentimental analysis has become one of the greatest innovation and efficient technique of information and web data analysis. Sentimental analysis refers to a classification problem where main focus is to predict the polarity of words and then categorize them into positive, negative and neutral feelings with the aim to identifying attitude and opinions that are expressed in any form or language. Proposed Recommendation System will analyze sentence and emoticon expressed and will discover proper mood based on which it will generate appropriate Status Suggestion which helps user in great way while expressing his/her emotions, E.g. If user has expressed Cool mood of emotion then proposed system will provide list of popular Status of Cool Emotion, that can be useful for user to post it as by its Emotion. This way the proposed system targets user's original behavior over social media. Also it can be useful as in field of marketing application as well as behavioral application too.

Keywords- **Text Classification, Sentiment Analysis, Social Media, Emotional Words, Emoticons, Recommendation System.**
_____

I. INTRODUCTION

In the past few years, micro-blogging platforms, such as twitter, are becoming most popular online social networks. Different opinions and news can be shared about various aspects and occasions using these micro-blogging platforms. Twitter is therefore considered as a rich source of data and it can be used for different text analysis and decision making tasks. The main focus of sentiment analysis is about text classification into positive/negative/neutral feelings based on the polarity of text.

Sentiment analysis refers to a broad area of natural language processing, computational linguistics, and text mining. Typically, the goal is to determine the polarity of natural language texts. An intuitive approach would involve scanning a text for cues signaling its polarity. In face-to-face communication, sentiment can often be deduced from visual cues like smiling. However, in plaintext computer-mediated communication, such visual cues are lost. Over the years, people have embraced the usage of so-called emoticons as an alternative to face-to-face visual cues in computer-mediated communication like virtual utterances of opinions. In this light, we define emoticons as visual cues used in texts to replace normal visual cues like smiling to express, stress, or disambiguate one's sentiment. Emoticons are typically made up of typographical symbols such as \:", \=", \-", \)", or \(" and commonly represent facial expressions. Emoticons can be read either sideways, like :-(" (a sad face), or normally, like " \(^ ^)" (a happy face).

TABLE I. SAMPLE TWITTER POSTS

Interacted with a delegation of farmers. Discussed various issues relating to agriculture & farmer welfare. :^)
Condemn the attacks in Ivory Coast & Ankara. My thoughts are with the families of the deceased. :'-(May those injured recover quickly..
Looking forward to meeting you all tomorrow. Do tune in. #MannKiBaat.. :D
Delighted to join program at Patna HC. Congrats to HC for completing a century & best wishes in serving the people in the years to come.. :)

II. RELATED WORK

Many researchers have proposed different studies in recent years in order to classify the sentiments expressed in a tweet as positive, negative or neutral. Some of the research work is overviewed in the next few paragraphs.

Twitter streaming API provides twitter feeds which are further refined using different preprocessing and text analysis tasks. Then refined tweets are classified using emoticon classifier, Bag of words classifier and Sentiwordnet classifier sequentially. The proposed algorithm resolves limitations of existing algorithms and increases classification accuracy with great ease. has been proposed by F.H.Khan [1].

Hogenboom et al [2] in 2013, on this corpus, paragraph-level accounting for sentiment implied by emoticons significantly improves sentiment classification accuracy. This indicates that whenever emoticons are used, their associated sentiment dominates the sentiment conveyed by textual cues and forms a good proxy for intended sentiment.

SHO AOKI et al [3] in 2011, it can be used to automatically generate the emotional vector of various emoticons using the emotional words that co-occur with such emoticons. Furthermore, we conducted an experiment collecting and examining a large volume of weblog articles, and showed the effectiveness of the proposed method.

III. PROPOSED METHODOLOGY

STEP 1: TWEETER DATASET:

Twitter users generate about 200 million tweets (short messages of up to 140 characters) per day. Here using import.io API extract tweets.

STEP 2: PREPROCESSING:

This proposed technique contains different preprocessing steps to make data more efficient before classification. Main Data have raw information like Emoticons, hash tags, URLs, special characters, abbreviations, slangs & stop words. Such data must be processed in arrange to maintain quality analysis. Using spelling correction with dictionaries such as, WordNet, SpellCheck and JSpell. For abbreviations or slangs, Netlingo and sms dictionaries are used. [1]

Preprocessing includes main three parts like:

A] Sentence Splitting
In sentence splitting just split paragraph into sentence from end words.

B] Tokenization
Tokenization method divides the text of a sentence into sequence of tokens and creates results in tokens consisting of one single word (unigram).

C] POS tagging
Last is the process of Part-Of-Speech tagging (POS) allows to automatically tag each word of text in terms of which part of speech it belongs to:
Adverb, noun, pronoun, adjective, verb, interjection, intensifier etc.

D] Words Polarity Dictionary
Word of dictionary contain list of positive negative and neutral words.
In this dissertation, three main parts further divided into ten parts which contain synonymous or relevant words list:

Positive includes:
Happy, Trust and Love

Negative includes:
Sad, Fear, Anger, Disgust and Awe.

Neutral includes:
Surprise and Anticipation.

IV. EVALUATION

- ➢ Each dataset contain Total Tweets Sentences = 50

CALCULATE ACCURACY FOR FIRST DATASET Narendra Modi:

- ➢ From System Correctly Classified Sentences are : 42
- ➢ From System Incorrectly Classified Sentences are : 8
- ➢ System Overall Accuracy = 42/50 = 84 %

Also same procedure applied on other dataset and based on analysis various statics have been mentioned in below table:
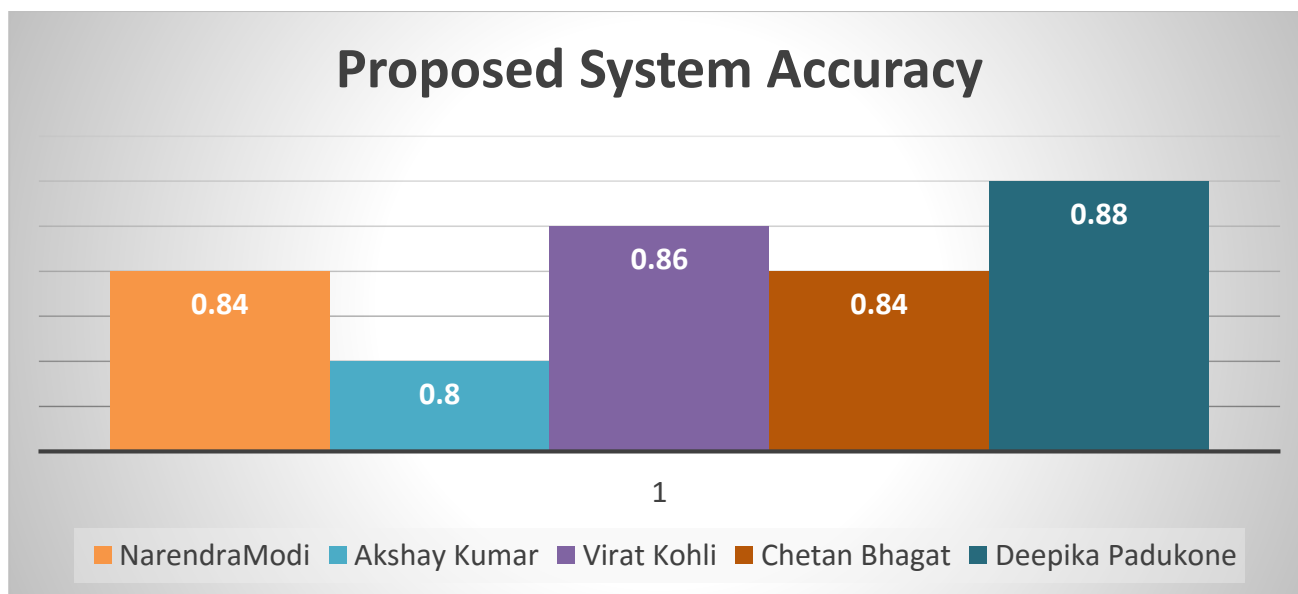

Figure 1: Proposed System Accuracy over five different dataset

V. CONCLUSION & FUTURE WORK

This Dissertation targets to combine smiley and statement, both, to improve result of classification of user's Emotion properly from Social Media Post. According to that, the target system combined with recommendation system and through classification of old posts of user; the system provides user the next list of status suggestion.

- •   The proposed system generates more accurate and efficient sentimental analysis with using emoticons.

- •   This sentiment analysis will be very helpful for status suggestion according to their proper mood and use that status in various social media.

FUTURE WORK

- •   Design more efficient GUI and integrating it for any social media applications.

- •   Increase size of sentiment word dictionary as well as emoticon dictionary for more efficient result.

VI. REFERENCES

1. F.H.Khan,U.qamar, M.Y.Javed, "SentiView: A visual sentiment analysis framework", NUST ,978-1-908320-38/4/2014, IEEE .
2. A.Hogenboom, D.bal,F.Frasincar, "Exploiting Emoticons in Sentiment Analysis", SAC'13 March 18   22,2013, Coimbra, Portugal.
3. SHO AOKI,OSAMU UCHIDA,"A Method for Automatically Generating the Emotional Vectors of Emoticons Using Weblog Articles",ISBN:978-960-474-281-3,2011.
4. D.Terrana,A.augello,G.Pilato, "AUTOMATIC Unsupervised Polarity Detection on a Twitter Data Stream",978-1-4799-4003-5,ICSC,2014   IEEE.
5. S.Huang,W.han,X.Que, W.Wang, "Polarity Identification of Sentiment Words based on Emoticons",978-1-4799-2548-3, CIS,2013 IEEE.
6. Jichang Zhao, Li Dong, JunjieWu, and Ke Xu, "MoodLens: An Emoticon-Based Sentiment Analysis System for Chinese Tweets", ACM 978-1-4503-1462-6, KDD'12, August 12–16, 2012 ACM.
7. M.Boia,B.Faltings,C.musat,P.Pu,"A :) Is Worth a Thousand Words: How People Attach Sentiment to Emoticons and Words in Tweets", 978-0-7695-5137-1; 2013, IEEE.
8. Lun-Wei Ku, Cheng-Wei Sun, "Calculating Emotional Score of Words for User Emotion Detection in Messenger Logs", 978-1-4673-2284-3,2012 IRI IEEE.
9. Chetan Mahajan,Preeti Mulay, "E3:Effective Emoticon Extractor for Behavior Analysis from Social Media", ISBCC'15,2015   Science   Direct.
10. Walaa Medhat ,Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithm and applications: A Survey ", 2090-4479,2014, Ain Shams Engineering Journal, Elsevier.
11. Lumin Zhang, Yan Jia, Bin Zhou, Yi Han, "Micro blogging Sentiment Analysis Using Emotional Vector", 978-0-7695-4864-7, 2012, Second International Conference on Cloud and Green Computing, IEEE, Computer Society.
12. Hao Wang, Jorge A. Castanon, San Jose, "Sentiment Expression via Emoticons on social Media", 2015, IEEE.
13. J.S.guerrero, Jose A. olivas ,F.P.Romero,Sentiment Analysis :A Review and Comparative Analysis of Web Services", department of information technologies and systems, March 19,2015 Elsevier.