ARTIFICIAL EMOTIONAL INTELLIGENCE ON TWITTER USING INFORMATION EXTRACTION AND NATURAL LANGUAGE PROCESSING

M. Navaneethakrishnan, Professor of Computer Science & Engineering,
G. R. Sandeep, Student of Computer Science & Engineering,
S. Santha Kumar, Student of Computer Science & Engineering,
St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract

Social networking sites are databases with a lot of information. Twitter, for example, generates millions of bytes of data. These records can be utilized for both commercial and social purposes. One of the new buzzwords for many corporate strategies is analyzing data from these social networking Web sites. Sentiment analysis may be used to efficiently manage election campaigns, global health crises, technical concepts, innovations, entertainment, and natural resources.

Introduction

Sentiment analysis is a well-known example of text-based affect computing. This domain's previous work has focused on corpus building, lexical or syntactic feature extraction from the corpus, and ultimately close testing, which assigns a polarity label to each document or other text unit in the corpus. More and more social networks are now allowing sociologists to do analyses that they previously couldn't. In this research, one of the most well-known websites, Twitter, which provides social networking capability, is used to see if the environment that users are exposed to in the online social realm has an impact on their emotions. We create models to learn both emotional influencing and influenced probabilities for users, and we report our findings using actual social network data.

Facebook, Twitter, and WhatsApp have all become popular in recent years. The opinions, news, and conversations available on these platforms aid in the prediction of results. These are the places where one may express oneself on the internet. Twitter is one such network that gets popular in our country following major events. On Twitter, everything that is news becomes a trend. As a result, this project examines tweets gathered from Twitter and applies emotive analysis to them.

Literature Survey

COVID Senti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis was proposed by Usman Naseem, Imran Razzak, Matloob Khushi, Peter W. Eklund, and Jinman Kim. It entails gathering all user-provided tweets about Corona, Coronavirus, Covid, Covid-19, sickness, and related topics, as well as offering the user's opinion or mood.

Sai Madhu, Chakradhar Reddy and Polireddy M proposed Sentimental Analysis on

Twitter. It collects all types of tweets and analyzes them and produces the polarity of the tweets as positive, negative and neutral.

Hamid and Johirul demonstrated the use of sentiment analysis for Twitter data. They looked at emotion mining for a variety of issues, from politics to humanism, and came to a conclusion about its impact.

Predicting Stock Price using Sentimental analysis through Twitter Data was proposed by Niveditha N Reddy, Naresh E, Vijaya Kumar B P which predicts stock price using Twitter tweets.

Mathematical Representation

TF-IDF

a) It works on the principle-based grouping of most common or repeating words in a particular document. It divides each very important word and tries to make calculations based on the following given formula.

b) Once the data is labeled, implement the TF-IDF before implementing K means clustering, and it is required to create TF and IDF matrices separately.

c) Term Frequency–Inverse Document Frequency (TFIDF) is defined as a multiplication of two Matrices, which are TF and IDF, and informs the significance of a word in a sentence or statement or a document.

Formula for **TF-IDF= TF*IDF**

Where, TF= (Number of repetition of words in a given or particular sentence) / (Number of words in the given sentence)

Where, IDF=log (No of sentences / (No of sentences contains the word)

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

p, q = two point in Euclidean n-space

qi, pi = Euclidean vectors starting from the origin of the space(initial point)

n = n-space

System Design

The suggested system gathers all types of tweets from users and performs sentiment analysis on them. This system captures every tweet based on the keywords entered by the user by using Twitter API. The tweets are processed using Natural Language Processing (NLP), and the

polarity of the tweets is classified as Positive, Negative, or Neutral.

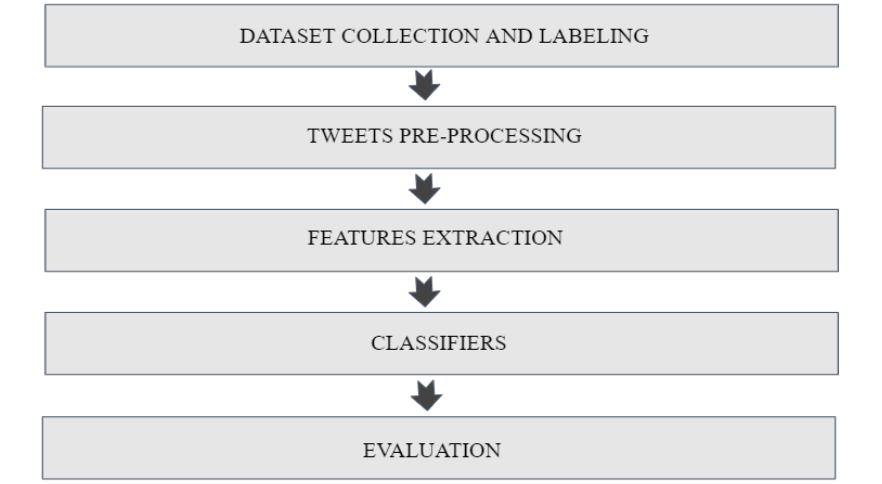

Fig 1. Steps involved in Twitter Sentimental Analysis

Algorithm

Input: Unlabeled Tweet: Tunlabeled,
Output: Labeled Tweet: Tlabeled,
Compute:
Positive: Tpositive = [];
Negative: Tnegative = [];
Neutral: Tneutral = [];
Steps:
for t in T(t) do:
If (t is English):
Perform Labeling: (TextBlob)
If (-1 < polarity of t < -0.4):
Negatively labeled
If (0.4 < polarity of t < 1):
Positively labeled
else:
Neutrally labeled
Perform Pre-processing:
Remove punctuation and stop-words, perform stemming, and convert tweet to lower-case
else:
Delete t
end for
Output:
Pre-processed labeled Tweets: Tl = [Tpositive, Tnegative, Tneutral]

Implementation

Dataset Collection and Labeling

Multiple datasets can be combined into a single object using dataset collections. As a result, instead of dealing with four, a hundred, or a thousand different datasets, you only have to deal with one item in Galaxy history. We need to establish a hierarchical collection named Coupled Dataset Collection or Paired Collection since our data is paired.

Tweets Preprocessing

One of the most important processes or techniques is data cleaning. It entails deleting any data that is no longer needed. The Hive and Google Colabs databases are used to obtain the data. And then the pre-processing begins when the dataset is imported.

Let's have a look. "@User1:I am a decent guy," says the phrase. If you look at this statement, you'll notice that the '@' sign must be removed, as well as User1.name. Because the user's name is User1, it has been deleted. deleting an entry that previously existed in the name filed column superfluous tags such as hyphens, Maastricht (*), and the like, as well as the "I am a decent guy," the phrase will change to.

Features Extraction

Feature extraction is a dimensionality reduction procedure that reduces a large set of raw data into smaller groupings for processing. The enormous number of variables in these large data sets necessitates a lot of computational resources to process.

Classifiers

In machine learning, a classifier is an algorithm that automatically sorts or categorizes data into one or more "classes." One of the most prominent instances is an email classifier, which examines emails and filters them according to whether they are positive, negative or neutral.

Evaluation

It can be inferred that greater work was put into doing the inquiry on such a large amount of data as well as analyzing the tweets to anticipate the cluster's quality, whether good, negative, or neutral. Subjectivity is used to determine if a word is subjective or objective. However, polarity informs us about human behavior, both good and negative, for each term. There is a point "zero" in the polarity column. As a result, all points on the graph to the left of zero represent negative remarks, while those to the right of zero represent good comments.

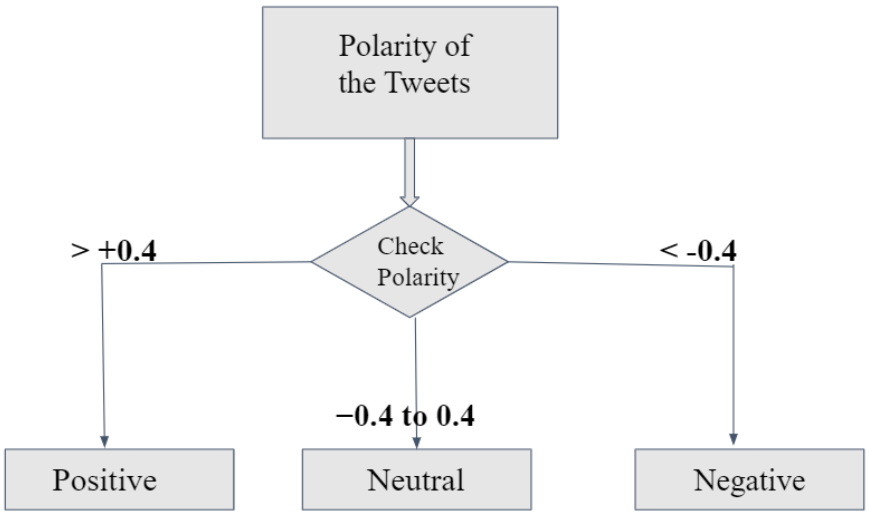

Fig 2. Polarity Checking

Fig 3. Wordcloud

Conclusion and Future Work

It can be inferred from the suggested study that more effort was put into doing the inquiry on a large volume of data and also analyzing the tweets to anticipate the cluster's quality, whether good, negative, or neutral. For analysis, a variety of algorithms are utilized, including VADER, which is a programme that describes how tweets are projected to be good, negative, or neutral. To produce more accurate findings, the K-Means clustering strategy is used with the TF-IDF technique.

References

[1] Usman Naseem, Imran Razzak, Matloob Khushi, Peter W. Eklund, and Jinman Kim. "COVID Senti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis."(2021 - 2022)

[2] Sai Madhu, Chakradhar Reddy and Polireddy M. "Real Time Sentiment Analysis on Twitter. "(2021)

[3] Hamid and Johirul. "The use of sentimental analysis for Twitter Data. " (2018)

[4] Niveditha N Reddy, Naresh E and Vijaya Kumar B P. "Predicting Stock Price using Sentimental analysis through Twitter Data." (2021)

[5] Salloum, Said A., et al. "A survey of text mining in social media: facebook and twitter perspectives." Adv. Sci. Technol. Eng. Syst. J 2.1 (2017): 127-133.

[6] Balazs, Jorge A., and Juan D. Velásquez. "Opinion mining and information fusion: a survey." Information Fusion 27 (2016): 95 -110.

[7] Devi, K. Nirmala, and V. Murali Bhaskarn. "Online forums hotspot prediction based on sentiment analysis." (2012).

[8] Kannangara, Sandeepa. "Mining twitter for fine-grained political opinion polarity classification, ideology detection and sarcasm detection." Proceedings of the Eleventh ACM International Conference on Web Search and Dat a Mining. 2018..

[9] Gull, Ratab, et al. "Pre processing of twitter's data for opinion mining in political context." Procedia Computer Science 96 (2016): 1560- 1570.

[10] D. Chakrabarti and K. Punera, “Event summarization using tweets,” in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.

Biography

Dr.M.Navaneethakrishnan M.E., PhD is a Head of the Department in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. He completed his Ph.D, in Cyber Security - Computer Science and Engineering in 2017 from Manonmaniam Sundaranar University (MSU) Tirunelveli, Tamilnadu. He has done his M.E, CSE in Anna University Chennai in the year 2008. Dr.M.Navaneethakrishnan has 15 years of teaching experience and has 58 publications in International Journals and Conferences. His research interests include network security, Computer Networks, data science and Machine Learning. He is an active member of ISTE, CSI, IEANG and IEI.

Mr. G.R.Sandeep B.E., Computer Science and Engineering at St.Joseph College of Engineering, Sriperumbudur, Chennai, Tamilnadu. I have attended many Conferences and Workshops on Machine Learning and Artificial Intelligence. I have also done the Certification Course on Machine Learning and Python which was conducted by Indian Institute of Technology, Kanpur and E & ICT Academy.

Mr. S. Santha Kumar B.E.,Computer Science and Engineering at St.Joseph College of Engineering, Sriperumbudur, Chennai, Tamilnadu. I have attended many Conferences and Workshops on Machine Learning and Artificial Intelligence. I have interests in Machine Learning, Cloud Computing, and App Development..