School of Engineering and Applied Science (SEAS), Ahmedabad University

B.Tech (CSE Semester VI)/M.Tech/PhD:
Machine Learning (CSE 523)

Project Abstract Submission #1

Submission Deadline: January 26, 2020 (11:59 PM)

- **Group No.: B-NLP5**
- **Project Area: Natural Language Processing**
- **Project Title: Harnessing Twitter 'Big Data' for Automatic Emotion Identification**
- **Name of the group members :**
  1. Jay Patel (1741018)
  2. Manav Shah (1741042)
  3. Prima Sanghvi (1741045)
  4. Priyanshi Deliwala(1741047)

Abstract

Introduction

Natural Language Processing is a technology which trains computer to understand how human communicate. The ultimate objective of the problem chosen-emotion identification, is useful to gain public opinion concerning particular topics. It is the process of determining the emotional tone behind a series of words expressed on online platform.

The application of emotion identification is very broad and influential. For example, it helps to quickly understand the consumer behaviour or attitudes and react accordingly to take an advantage.

Background and Motivation

It is as yet hard for a greater part of devices to definitely assess what genuinely is a negative, neutral, and a positive explanation. As dealing with sarcasm or context behind the script is not dealt successfully, it gets difficult to understand the mechanics behind the same.

User generated content on twitter provides a rich source for predicting people's emotion, which is

in turn necessary and important for deeper comprehension for people's actions.[1] Due to relatively small training datasets studies on emotion identification, lack apprehensive content of emotional situations. To overcome this bottleneck, we have created a large emotional-based dataset/content. Learning techniques in order to improve emotion identification in other domains.

Contribution

After providing a reliable emotion identification mechanism with the help of 2.5 million tweets' database which comprises of execution of two of the most efficient algorithms, LIBLINEAR and MNB(Multinomial naive bayes) we acquire a reliable system for emotion identification. By this project, we demonstrate that a combination of uni-grams, bi-grams, sentiment/emotion-bearing words, and parts-of-speech information is most effective for gleaning emotions.[2] As this article purely focuses on classification, we examine the accuracy of the model using varied size dataset. We use this project as a base to understand natural language processing and machine learning more profoundly.
Feature reduction is done to make the classification more accurate and efficient as online platform datasets are quite large.[3] Principal component analysis is linear method, which we will be using to reduce the dimensionality of dataset, as it results in small amount of loss in variance.
Regression is a supervised learning technique is used to detect the weightage of emotion in the range of 0 to 1. We would train a regression model to predict the sentiment polarity and intensity of words or phrases in the tweet, by splitting the dataset into training and testing sets.[4]
Kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable.The main idea of this algorithm is to obtain the probability density function of the original text data by kernel density estimation method.[5] Thus the original training set can be expanded and the traditional classifiers can obtain a better classification performance.

References

[1] M. D. Choudhury, S. Counts, and M. Gamon, "Not all moods are created equal! exploring human emotional states in social media," in Proceedings of ICWSM, 2012.

[2] G. Mishne, "Experiments with mood classification in blog posts," in Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access

[3] Gour, Rinu. "Dimensionality Reduction in Machine Learning." Medium, Medium, 8 Apr. 2019, medium.com/@rinu.gour123/dimensionality-reduction-in-machine-learning-dad03dd46a9e.

[4] Suhasini, Matla Badugu, Srinivasu. (2018). Two Step Approach for Emotion Detection on Twitter Data. International Journal of Computer Applications.

[5] Yang, L. [online] Pdfs.semanticscholar.org. Available at: https://pdfs.semanticscholar.org/ce34/4a782fd36fdaf7 [Accessed 31 Jan. 2020].