Natural Language Processing

UML602

Lab Project

Submitted by:

Chirag Mahawar 101603078

Tanish Charaya 101611056

COE-6

Submitted to :

Ms. Akashdeep Kaur

Computer Science Department

Thapar Institute of Engineering & Technology, Patiala

May 2019

CONTENTS

INTRODUCTION

Also known as “Opinion Mining”, Sentiment Analysis refers to the use of Natural Language Processing to determine the attitude, opinions and emotions of a speaker, writer, or other subject within an online mention.

Essentially, it is the process of determining whether a piece of writing is positive or negative. This is also called the Polarity of the content. As humans, we are able to classify text into positive/negative subconsciously. For example, the sentence “The kid had a gorgeous smile on his face”, will most likely give us a positive sentiment. In layman’s terms, we kind of arrive to such conclusion by examining the words and averaging out the positives and the negatives. For instance, the words “gorgeous” and “smile” are more likely to be positive, while words like “the”, “kid” and “face” are really neutral. Therefore, the overall sentiment of the sentence is likely to be positive.

A common use for this technology comes from its deployment in the social media space to discover how people feel about certain topics, particularly through users’ word-of-mouth in textual posts, or in the context of Tweets, App reviews, Restaurant reviews etc.

Naive Bayes, Decision Tree, Random Forest Classifiers are widely used classification models in NLP. Here we have used Random Forest Classifier.

Random Forest Classifier

Random Forest is a supervised learning algorithm. Like you can already see from its name, it creates a forest and makes it somehow random. The forest it builds, is an ensemble of Decision Trees, most of the time trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. To say it in simple words: Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is, that it can be used for both classification and regression problems, which form the majority of current machine learning systems.

STEPS OF WORKING

- First we read the csv file which has all the reviews and the corresponding sentiment label(Positive, Negative, Neutral).
- Preprocessing of the dataset, which includes removal of numeric characters and punctuations and stopwords and reviews of length less than 1.
- Now preprocessed dataset is stemmed and a preprocessed corpus is generated.
- Now using label encoder labels are transformed to numeric values.
- Positive -> 2
- Neutral -> 1
- Negative -> 0
- Now for each review using countVectorizer the np array of dataset is created.
- Training the model using RandomForestClassifier
- Visualizing the results

APPLICATION

Sentiment analysis is a technique which allows big companies to categorize the massive amount of unstructured feedback data from social media platforms. Following are some ways in which this technique is applied to actual feedback monitoring:

1. Finding hot keywords:

Opinion mining can majorly help in discovering hot search keywords. This feature can help the brand in their SEO (Search Engine Optimization). This means that opinion mining will help them make strategies about, how their brand will come up among the top results, when a trending or hot keyword is searched in a search Engine.

2. Voice of customer:

Sentiment analysis of social media reviews, mentions and surveys help to broadcast the voice of customers to the brand, they are expressing their views about. This way the brand knows, exactly how common folk feels about their services. The company can use this information in growing their market, advertisement targeting and building loyalty among its customers.

3. Employee feedback:

Sentimental analysis can also be used to receive feedback from the employees of the company and analyze their emotions and attitude towards their job. And to determine whether they are satisfied with their job or not.

4. Better services:

Text mining can provide a filter about, which service of the company is getting more negative feedback. This will help the company to know, what are the problems arising with that particular service. And based on this information the company can rectify these problems.

5. Get to know what's trending:

This will not only help the company to stay updated and connect more with the audience, but it will also facilitate the rise of new ideas, for developing new products. This will allow the company determine what the majority of the audience demands, and develop a product according to these demands.

6. Feedback on pilot releases and beta versions:

When a company releases a new product or service, it is released as a pilot or beta version. The monitoring of public feedback at this stage is very crucial. So, text mining from social media platforms and review sections greatly helps accelerate this process.