2021

Arabic Sentiment Analysis

PROPASL FOR PROJECT 1

KHADEJAA SAAD MOHAMMED ALSHEHRI

Sentiment analysis is a popular and challenging Natural Language Processing (NLP) task, which facilitates capturing public opinions on a topic, product or a service. It is thus a widely studied research area. In this project, we will try to classify sentiment analysis for Arabic tweets into three classes which are positive, negative and neutral.

Question/need:

- How many tweets we have in each class?
- How many unique words for each class?
- What is maximum number of words for the longest tweet in each class? And what is the mean and average of the tweets long ?
- What is the common words used for each class?
- What is the common emoji used in each class?

Data Description:

In this project, We will use ASAD public dataset which is intended to accelerate research in Arabic NLP in general, and Arabic sentiment classification in specific. ASAD is publish by KAUST for sentiment analysis computation last year and publish for public used this month in Kaggle website[i].
The individual sample in our dataset is tweet with it's label (positive, negative or neutral). So, we expect to work with textual features using pre-trained model such as tf-idf, word2vec ..etc. Also, we plan to use ML algorithm such as SVM and DL model.

Tools:

- Pandas
- Numpy
- Sklearn
- Keras
- And others

[i] https://www.kaggle.com/c/arabic-sentiment-analysis-2021-kaust/data