EmoInt

Nithin S K

***Abstract*—The paper describes a system for EmoInt - a shared task to predict the intensity of emotions in tweets. . The emotions are classified as - anger, fear, joy and sadness, whose intensities have real valued score of between 0 and 1.**
Two methods are presented here - one based on SVM regression model and another based on a deep learning model.

***Index Terms*—EmoInt.**

I. Introduction

EMOINT (Mohammad and Bravo-Marquez, 2017) is a shared task hosted by WASSA 2017, aiming to predict the emotion intensity in tweets. The emotion of each tweet can be any one out of anger, joy, fear and sadness. The task is to predict the emotion intensity of the ,for each tweet whose emotion is known .The intensity is a real valued score ranging from 0 to 1.

II. Data

Different dataset are provided for train, dev and test. In each of these dataset, there are four collection , each belong to the the emotions under consideration.

III. Proposed System

Two different approaches are explored for the task. One based on an SVM regression model and another based on a deep learning model

A. Input Features

Text from each tweet is cleaned by removing links ,stop-words, , numbers and transforming all alphabets to lower case. The words are tokenized and Tfidf features are extracted from them.
TF-IDF (term frequency-inverse document frequency) is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. This is done by multiplying two metrics: how many times a word appears in a document, and the inverse document frequency of the word across a set of documents.
For each tweet a feature vector of size 1000 is extracted. Categorical labels are assigned to each of the four emotions. This label is also concatenated to the feature vector. Finally, for each tweet a 1004 long feature vector is obtained.

B. SVM Regresion model

The SVR implementation of Support Vector Regression from scikitlearn is employed here.Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of points.

C. Deep learning model

Deep learning models are built using neural networks. A neural network takes in inputs, which are then processed in hidden layers using weights that are adjusted during training. Then the model gives out a prediction. The weights are adjusted to find patterns in order to make better predictions.
Deep learning model with four hidden layers is employed.The first three dense layers have sizes 1004, 512 and 64 with reLu activation. Dropout layers with 0.2 as parameter are provided between them. Adam optimizer is used and Mean Average Error as the loss. Sigmoid activation is given for the output layer.

IV. Results

Both models are trained and prediction obtained for the test data. The test data consist of different collection of tweets having different emotions. Seperate models are used for anger, sadness,fear and joy. Out of the employed model the SVM regression performed better. As per the evaluation criterion of the task, the average Pearson correlation for SVR is 0.53 and that for the Deep learning model is 0.044.

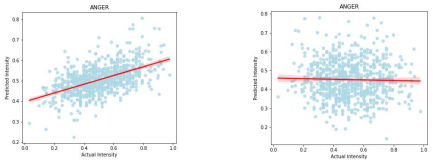

Fig. 1. SVR vs Deep model ANGER

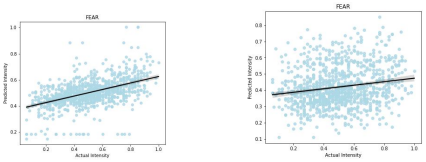

Fig. 2. SVR vs Deep model FEAR

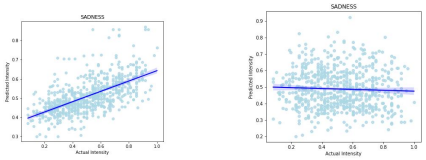

Fig. 3. SVR vs Deep model SAD

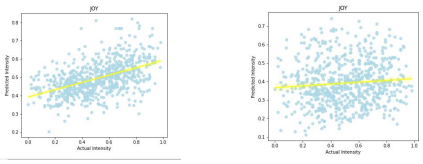

Fig. 4. SVR vs Deep model JOY

REFERENCES

[1] WASSA-2017 Shared Task on Emotion Intensity. Saif M. Mohammad and Felipe Bravo-Marquez. In Proceedings of the EMNLP 2017 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA), September 2017, Copenhagen, Denmark.