Text Correlation with EMOJI

TA Mentor : Satyam Mittal

Team Members : Akshat Maheshwari, Kalpit Pokra,

Kushagra Nagori, Kartik Garg

Introduction

Problem statement

There are 2 parts to the problem statement of the project:

1. Given a set of tweets, classify the tweets according to the emojis.

2. Given a sentence, check the sentiment analysis and insert appropriate emojis in sentence.

First Step : Preparing Dataset

Datasets Used

We used 3 different datasets in order to obtain accurate results.

1. **Scrapped twitter data and prepared an annotated dataset.**
2. **A pre-annotated dataset with relatively larger range of emojis (around 500K tweets). (**Link**)**
3. **A clean dataset with proper emoji annotations for sentences. (**Link**)**

Problems : The need for multiple datasets?

1. Preprocessing of data.
2. Manual annotations for scraped data (first).
3. The first 2 datasets resulted required manual cleaning, resulted in a dirty dictionary/ bag of words.
4. The third dataset was relatively efficient and clean but was not scalable.

Emojis Analysed

Table of emojis, corresponding labels and the sentiment they portray.

code	emoji	label
:heart:	❤️	0
:baseball:	⚾	1
:smile:	😀	2
:disappointed:	😞	3
:fork_and_knife:	🍴	4

0	❤️	_red_heart_
1	😍	_smiling_face_with_hearteyes_
2	😂	_face_with_tears_of_joy_
3	💕	_two_hearts_
4	🔥	_fire_
5	😊	_smiling_face_with_smiling_eyes_
6	😎	_smiling_face_with_sunglasses_
7	✨	_sparkles_
8	💙	_blue_heart_
9	😘	_face_blowing_a_kiss_
10	📷	_camera_
11	🇺🇸	_United_States_
12	☀️	_sun_
13	💜	_purple_heart_
14	😉	_winking_face_
15	💯	_hundred_points_
16	😁	_beaming_face_with_smiling_eyes_
17	🎄	_Christmas_tree_
18	📸	_camera_with_flash_
19	😜	_winking_face_with_tongue_ :)

Tweet Classification Based on Emojis

Approach

1. Preprocessing
2. Tokenization
3. Bag of words
4. Feature Vector Generation
5. Model Implementation
6. Training and Validation
7. Testing Models

Models Implemented

1. Decision Tree
2. Linear SVM Classifier
3. Logistic Regression Classifier
4. LSTM Classifier

Results

MODEL	ACCURACY
Decision Tree	20.7%
Linear SVM Classifier	22%
Logistic Regression Classifier	26.7%
LSTM Classifier	38%

EMOJI PREDICTION FOR A SENTENCE USING SENTIMENT ANALYSIS

Approach

1. One-Hot Vector Creation
2. Loading word embedding vectors
3. Model Creation
4. Training and Validation
5. Testing

Models Implemented

1. Iterative Model (non-neural net model)
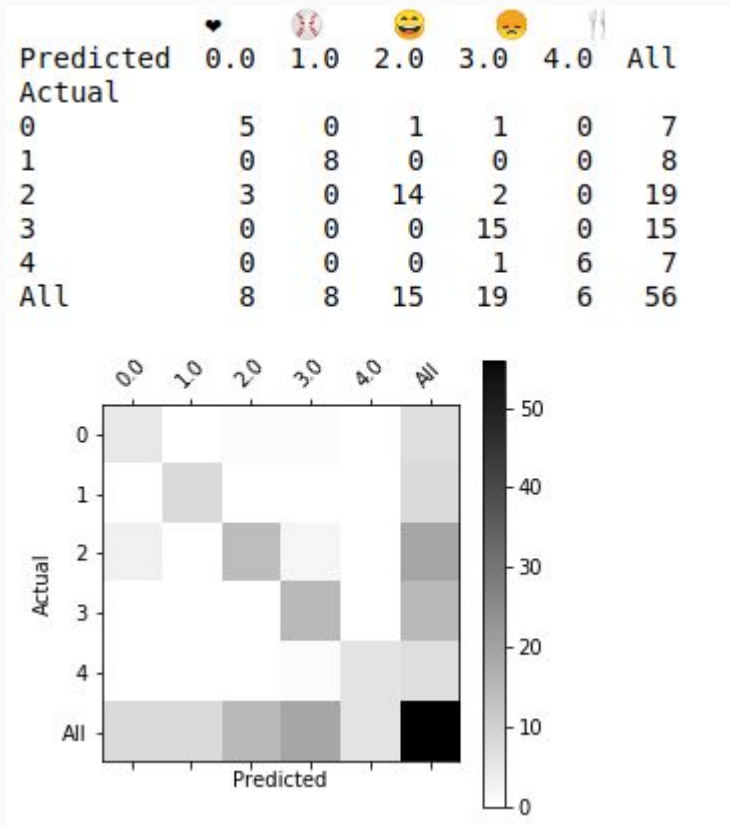2. LSTM Model

Results

- **Iterative Model**

  - Training Accuracy : 0.9587301587301588

  - Test Accuracy : 0.8071428571428571

- **LSTM Model**

  - Training Accuracy : 0.98 (approx)

  - Test Accuracy : 0.8171428571428571

Results for Iterative Model

References

The following research papers have been used:

1. Emoji Prediction in Tweets (Alison P. Ribeiro, N´adia F. F. da Silva)
2. Multimodal Emoji Prediction (Francesco Barbieri, Miguel Ballesteros, Francesco Ronzano, Horacio Saggion)

Questions?

THANK YOU :))