Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm

COMP8240 Group D Project Proposal

Ajit Kumar Chaudhary (Student id:- 46124942)
Avinash Panday (Student id:- 46139001)
Saroj Bogati (Student id:- 45842620)
Moiz Ahmed Khan (Student id:- 45785031)

Macquaire University

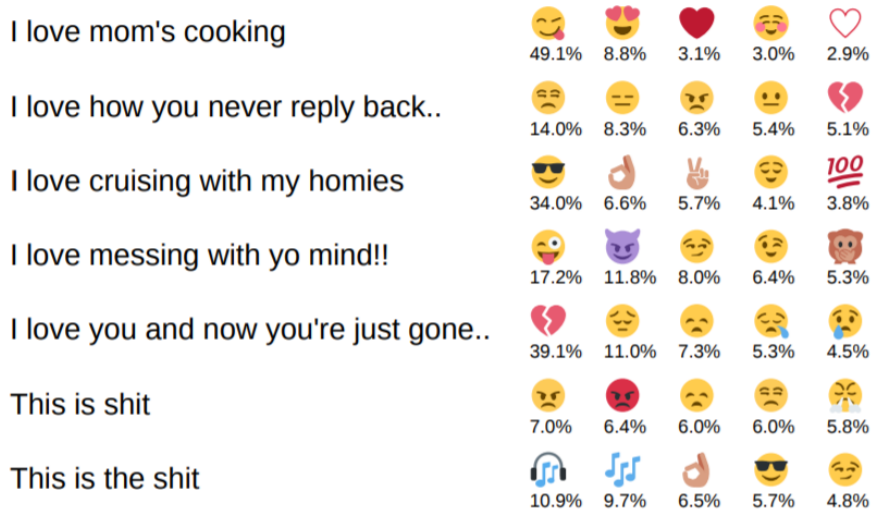
2020-09-01

Overview

- Motivation
- Benchmark datasets
- Method
- Demo

Motivation

It is published at **EMNLP** and globaly ranked 1st in **Sarcasm detection**. It uses 1246 million tweets containing about 64 common emojis on 8 benchmark dataset to predict sentiment, emotion, and sarcasm.



Benchmark datasets

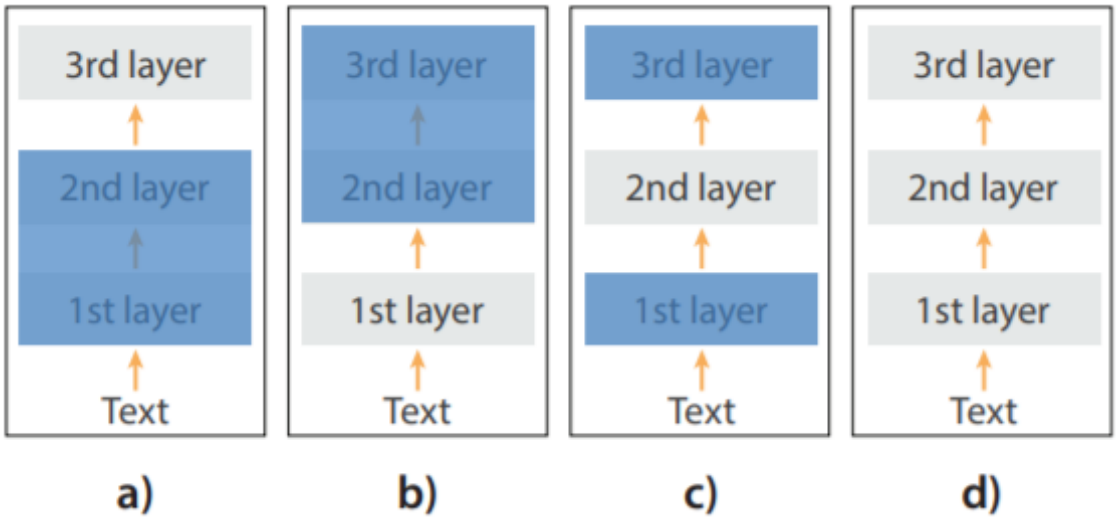
Code can be found in `https://paperswithcode.com/paper/using-millions-of-emoji-occurrences-to-learn`. It uses 8 benchmark datasets.

- **Olympic:** negative and high control, positive and high control, negative and low control, positive and high control.
- **PsychExp:** joy, fear, anger, sadness, disgust, shame, guilt
- **SCv1:** not sarcastic, sarcastic
- **SCv2-GEN:** not sarcastic, sarcastic
- **SE0714:** fear, joy, sadness
- **SS-Twitter:** negative, positive
- **SS-Youtube:** negative, positive
- **kaggle-insults:** neutral, insult

Method

- **Pretraining:-** The data is split into a training, validation and test set, where the validation and test set is randomly sampled in such a way that each emoji is equally represented.
- **Model Used:-** It uses LSTM, fast text classification, DeepMoji model to predict the accuracy of the classifiers.
- **Transfer learning:-** It uses approach in which all layers in the model are frozen when fine-tuning on the target task except the last layer.

Method



In the figure, each layer is fine tuned seperately. The layers covered in blue rectangle are frozen.

Demo

We will be reproducing the demo as per `https://deepmoji.mit.edu/`