Reproducing

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

Members:

Ajit Kumar Chaudhary

Avinash Pandey

Moiz Ahmed Khan

Saroj Bogati

Table of Contents

[Abstract 3](#_Toc55482190)

[Brief description of the source paper and justification 3](#_Toc55482191)

[Evaluation Framework 3](#_Toc55482192)

[Justification 4](#_Toc55482193)

[Description of original dataset 4](#_Toc55482194)

[Replication of original work 4](#_Toc55482195)

[Issues of original work and resolving those 5](#_Toc55482196)

[Construction of new data 5](#_Toc55482197)

[Twitter API 6](#_Toc55482198)

[Results on new data 6](#_Toc55482199)

[Any other reflections 6](#_Toc55482200)

[References 7](#_Toc55482201)

# Abstract

Reproducibility, a similar word to replicability, acts as a repetitive method of proving correctness, or reaching to similar suppositions as the expected results. When a published result is replicated with the usage of data, codes as well as papers delivered by the authors leading to expected results or similar, a research is said to be reproducible. Not just as a justification, but it also facilitates us as a revelation of research roadmaps as well as the transmission of knowledge. Here in this report, we will explain about our reproducible research based on a paper that used millions of emoji occurrences to study any domain illustrations in order to detect sentiment, emotion and sarcasm underlying.

# Brief description of the source paper and justification

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

## Evaluation Framework

## Justification

# Description of original dataset

In the paper, there are 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

# Replication of original work

The concept of empirical generalization greatly implies on any research being able to replicate. For our approach of replication, duplication or even extending the original work, there are sufficient tools available that facilitate us on the process. However, prior to the initiation, a good plan is needed that ease us with rectifying and evaluating necessary changes on the original work.

As per paper, we were able to successfully install and run source code using torchmoji model.

As per paper, there are following accuracy on the benchmark datasets.

We were able to replicate following accuracy on the benchmark datasets

using fine-tune of torchmoji model. Discussing about the model used here that is TorchMoji, which comes under DeepMoji Model. The TorchMoji model is being trained here along with various tweet that we fetched as a dataset. This model will help us to understand how language has a usage of expressing emotions. Likewise, with the usage of transfer learning, the model will help us to obtain state of art performance on many text modeling tasks related to emotions.

The original work consisted of emoji as well as score prediction. Emoji prediction consist of 5 most relevant emoji based on the sentence which you can see here. Similarly, Score prediction consists of accuracy score of the emoji in relevant to the emojis. Likewise, going through evaluation part, we had 8 benchmark datasets, and we could only replicate 3 datasets. Because it took approximately 12 hours for us to replicate those. Likewise, we made usage of 1000 epochs along with each epoch having sample size of 1000. The metrics used here is accuracy score along with labels. Similarly, we used fine tune model of torchmoji to facilitate us with the execution part which all comes under last method. So, talking about accuracy on these benchmark datasets that we executed, we got accuracy score of 31.95% in YouTube dataset, 41.20% in twitter dataset and 49.92% in Sarcasm dataset. Even though we tried evaluating for these 3 datasets, our main aim was to check and implement this model for twitter dataset, since we managed to create a new dataset from tweets fetched from twitter which will be explained further later part of the presentation. So, this brings to an end of explanation of original source code replication and evaluation part.

Note: - The model took too long to train (approximately 12 hours to train only 40% of the dataset.).

## Issues of original work and resolving those

# Construction of new data

Going through the dataset creation, we used twitter API and tweepy libraries to help us make a scrapping of tweets from the twitter. One of the prominent attribute of our dataset is that it consists of various tweet texts written by the same user. Here the data set will be used for making predictions of 5 most relevant emojis based on these sentences extracted from tweets. Also, we went through the process of data cleaning so that we get clean texts from raw tweets because there we many hashtags, mentions and other elements that we don’t in our dataset. Similarly, we also divided tweet texts into each sentences in order to get more specific results on relevant emojis

## Twitter API

We have used Tweepy library to extract data from twitter.

We have extracted 3200 tweets of Bill gates.

We have cleaned the text and generated labels with Positive and

Negative sentiments using TextBlob.

We have converted paragraphs of a tweet to sentences, therefore, we

have 5515 rows of tweet.

Evaluation of new data:

We have performed evaluation using finetune of torchmoji model.

We have received 92.20% accuracy on the new data.

# Results on new data

# Any other reflections

# References