

It's complicated 🎄

Matija Bertović, Antun Magdić, Ante Žužul

University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
autor1@xxx.hr, {autor2, autor3}@zzz.com

Abstract

Abstract...

1. Introduction

2. Related work

3. Dataset

4. Models

4.1. Model 1

4.2. Model 2

4.3. Bi-Directional LSTM

Recurrent neural networks currently have great results in modeling NLP tasks. For that reason, Long Short-Term Memory (LSTM) network is implemented. LSTM is a type of recurrent neural network that is able to capture long-term dependencies. Fully-connected layer is added after the LSTM cell to map the output vector size to the number of observed emojis. Each vector position in the output corresponds to a certain emoji, where the final output of the network is the emoji for which the output vector position is maximum, as given in formula (1).

$$\mathbf{y} = \underset{i}{\operatorname{argmax}} \mathbf{o} \quad (1)$$

Here, \mathbf{y} is the predicted emoji, \mathbf{o} is the output of the network and i is the position in vector \mathbf{o} . As there are 20 different emojis the model can predict, \mathbf{y} and i are integers ranging from 0 to 19.

Each input tweet is represented using pretrained word embeddings. Word embeddings are trained on twitter data using GloVe algorithm and are of size 100. Two bi-directional LSTM (BLSTM) layers with hidden size of 300 are used. Bi-Directional LSTM layer is composed of two LSTM layers, where one looks into the past and takes the sequence from the first word to the last and the other LSTM layer looks into the future and takes the sequence from the last word to the first. Both of those layers' outputs are then concatenated into one output of size 600. After the first BLSTM layer, dropout layer with dropout rate of 0.2 is used. In the end, fully-connected layer with size of 20 is used, because there are 20 different observed emojis. Parameters are optimized using ADAM algorithm with initial learning rate of 10^{-3} . Model is trained for 20 epochs over the train set with batch size of 32.

Table 1: Accuracy obtained on the test set for each model.

Model	Test accuracy
LR GloVe	33.78%
LR TF-IDF	53.35%
MLP GloVe	%
MLP TF-IDF	%
NB	%
kMeans	%
BLSTM	%

4.4. ...

5. Results

6. Clustering?

7. Conclusions