Results of Proposed Additions

Name Nikhil Meena
Roll No. 210667
Group C

Overview

Selected Paper: Automatic Depression Detection: an Emotional Audio-Textual Corpus and A GRU/BiLSTM-Based Model by Ying Shen, Huiyu Yang, and Lin Lin.

Integration of Transformer (BERT) for Textual Data Extraction

Implementation Details

The code which parses the dataset and converts it to a numpy array have been modified to give the text embeddings using 2 different transformer models : 'bert-base-chinese' and 'hfl/chinese-roberta-wwm-ext-large' separately.

Additionally, the training codes for bert and roberta have been added as well.

Results

After changing the text embeddings from ELMo to BERT, the text Bi-LSTM model converges much faster and gives more consistent results in the regression test along with some improvements in the MAE as mentioned below.

Text Embeddings	Text 1 - MAE	Text 2 - MAE	Text 3 - MAE
ELMO Embeddings (old)	8.42	8.23	7.64
BERT Tokenizer	7.94	7.60	7.39
Roberta	7.71	7.58	7.37

Table 1: MAE Results for Text Model using different embeddings

Text Embeddings	Fuse 1 - MAE	Fuse 2 - MAE	Fuse 3 - MAE
ELMO Embeddings (old)	7.82	8.24	7.52
BERT Tokenizer	7.34	8.22	7.93
Roberta	7.22	7.97	7.98

Table 2: MAE Results for Fusenet model with different. text models

Additionally, Initially with ELMo embeddings, the model results were not consistent and required multiple runs to get the best results on the particularly selected random indexes. But with Transformer embeddings for the text, the results are much more consistent and do not require multiple runs.

Training Time

The training time of the model is also faster with BERT embeddings. Since, the dimension of BERT embedding features are (162, 1, 768) whereas, the ELMo embeddings size is (162, 3, 1024) which is quite bigger.

Embeddings	Training Time (in min.)
ELMo	12:46
BERT	7:12

Table 3: Training times for different embeddings.

tested on Ryzen 7 5800U, 16GB RAM