Supplementary Material for ICON submission

Anonymous ICON submission

1 Hyperparameter setup for our analysis

All the models used the Adam optimizer and Softmax loss function. Values of the hyperparameters for each experiment are shown in Table 1.

1.1 Proposed approach: Fine tuning BERT and XLM-RoBERTa(XLM-R) models

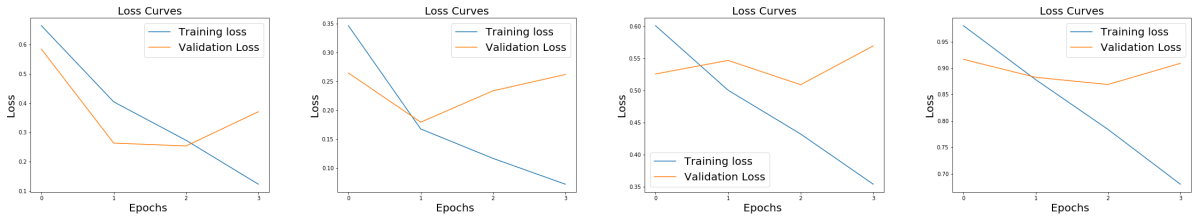
In our study, we have used two transformer models BERT and XLM-R, and fine-tuned them over the three different datasets. A Transformer reads entire sequences of tokens at once. In a sense, the model is non-directional, while LSTMs read sequentially (left-to-right or right-to-left). The attention mechanism in a transformer allows for learning contextual relations between words (e.g. 'his dog' in a sentence refers to 'John's dog').

BERT Tokenizer: For both the models we start by tokenizing the input sentences, pad/truncate the sentences to a fixed size, and then Create an array of 0s (pad token) and 1s (real token) called *attention mask*

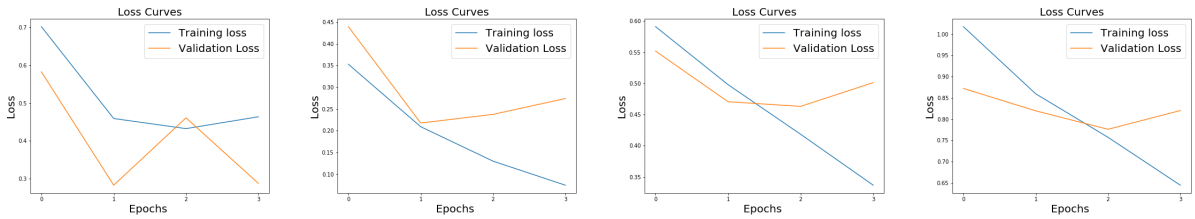
Training: Unlike (Islam et al., 2020), in our approach, we remove the CNN, GRU, and LSTM layers from the top of BERT/XLM-R and keep a fully connected layer for classification at the end. We set up a linear learning rate scheduler(*getlinearschedulewithwarmup*) for the warm-up of the layers during fine-tuning. Due to the linear scheduler, the learning rate gradually increases 0 to the initial learning rate set in the optimizer during the warm-up phase. After that, the learning rate will start to decrease linearly to 0. This leads to untrained classification layers to train more thereby bringing it up to the level of the pre-trained BERT layers. And then the learning rate reduces gradually.

Model	Train-Dataset	Test-Dataset	Learning Rate	Epochs	Batch Size
BERT+LSTM	PA(3-class)	PA(3-class)	5e-04	10	32
BERT+LSTM	PA(2-class)	PA(2-class)	5e-04	10	32
BERT+LSTM	YouTube-B	YouTube-B	5e-04	10	32
BERT+LSTM	Book-B	Book-B	5e-04	10	32
BERT+GRU	PA(3-class)	PA(3-class)	5e-04	10	32
BERT+GRU	PA(2-class)	PA(2-class)	5e-04	10	32
BERT+GRU	YouTube-B	YouTube-B	5e-04	10	32
BERT+GRU	Book-B	Book-B	5e-04	10	32
BERT+CNN	PA(3-class)	PA(3-class)	5e-04	10	32
BERT+CNN	PA(2-class)	PA(2-class)	5e-04	10	32
BERT+CNN	YouTube-B	YouTube-B	5e-04	10	32
BERT+CNN	Book-B	Book-B	5e-04	10	32
XLM-R+LSTM	PA(3-class)	PA(3-class)	5e-04	10	16
XLM-R+LSTM	PA(2-class)	PA(2-class)	5e-04	10	16
XLM-R+LSTM	YouTube-B	YouTube-B	5e-04	10	16
XLM-R+LSTM	Book-B	Book-B	5e-04	10	16
XLM-R+GRU	PA(3-class)	PA(3-class)	5e-04	10	16
XLM-R+GRU	PA(2-class)	PA(2-class)	5e-04	10	16
XLM-R+GRU	YouTube-B	YouTube-B	5e-04	10	16
XLM-R+GRU	Book-B	Book-B	5e-04	10	16
XLM-R+CNN	PA(3-class)	PA(3-class)	5e-04	10	16
XLM-R+CNN	PA(2-class)	PA(2-class)	5e-04	10	16
XLM-R+CNN	YouTube-B	YouTube-B	5e-04	10	16
XLM-R+CNN	Book-B	Book-B	5e-04	10	16
BERT-Fine	PA(3-class)	PA(3-class)	2e-05	4	16
BERT-Fine	PA(2-class)	PA(2-class)	2e-05	4	16
BERT-Fine	YouTube-B	YouTube-B	2e-05	4	16
BERT-Fine	Book-B	Book-B	2e-05	4	16
XLM-R-Fine	PA(3-class)	PA(3-class)	2e-05	4	16
XLM-R-Fine	PA(2-class)	PA(2-class)	2e-05	4	16
XLM-R-Fine	YouTube-B	YouTube-B	2e-05	4	16
XLM-R-Fine	Book-B	Book-B	2e-05	4	16

Table 1: Detailed setups of all applied experiments. BERT-Fine represents BERT fine-tuned model and XLM-R-Fine represents XLM-RoBERTa fine-tuned model. Note that, PA(3-class) represents Prothom Alo(3-class) dataset and PA(2-class) represents Prothom Alo(2-class) dataset.



(a) Book-B (b) YouTube-B (c) Prothom Alo (2-class) (d) Prothom Alo (3-class)



(e) Book-B (f) YouTube-B (g) Prothom Alo (2-class) (h) Prothom Alo (3-class)

Figure 1: Fig(a)-(d) shows training and validation loss curves for Fine-Tuning BERT model on Book-B, YouTube-B and Prothom Alo (both 2-class and 3-class) datasets. Fig(e)-(h) shows training and validation loss curves for Fine-Tuning XLM-RoBERTa model on Book-B, YouTube-B and Prothom Alo (both 2-class and 3-class) datasets.

References

Khondoker Ittehadul Islam, Md Saiful Islam, and Md Ruhul Amin. 2020. Sentiment analysis in bengali via transfer learning using multi-lingual bert. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, pages 1–5. IEEE.