

CS380 – ARTIFICIAL INTELLIGENCE LAB

Assignment-1: Text Classification

(Read all the instructions carefully & adhere to them.)

Date: 15th Jan 2024

Time: 3 to 6 PM

Total Credit: 20 (Implementation: 10, Findings/Observation Summarization: 5, One step ahead than requirement: 5)

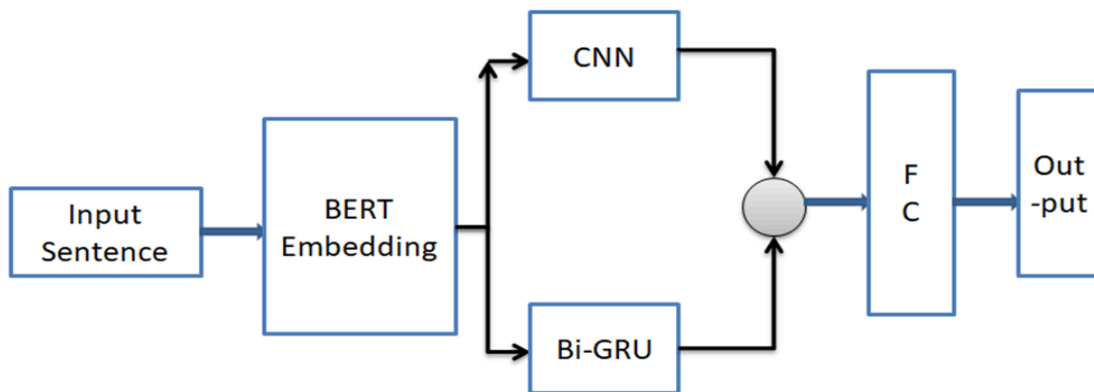
Instructions:

- Markings will be based on the correctness and soundness of the efficacy of the developed model.
- Comments/explanations/intuitions should be provided in a separate pdf file etc., document and not the code file.
- You are allowed to use existing frameworks and APIs. You can use either TF, Keras or Pytorch (recommended).
- Raw Dataset Link: [Hate Speech Dataset](#), Train, Valid and Test: 70, 10, 20, random seed: 08012023
- You can make the necessary assumption, please mention it in the report.

Problem:

Please go through the links below to get the Hate speech dataset. You have to implement a deep learning-based text classification model, where the BERT language model generates embedding (Max sequence length * 768) of the input sentence, which is passed through two channels. Channel-1 is BERT+CNN and channel-2 is BERT+biGRU. Finally, you have to concatenate the outputs returned by both channels. Then this concatenated feature will be passed through a Fully connected layer (**FC**) followed by a softmax output layer.

An overview of the above-mentioned deep learning model:



Datasets: Twitter Hate Speech dataset ([Hate_Speech_Dataset](#))

Class labels: 0 - hate speech, 1 - offensive language, 2 - neither

Evaluation Metrics

- (a) Accuracy, Precision, Recall and F1
- (b) Confusion Matrix
- (c) Training Time and Inference Time