Homework 1B

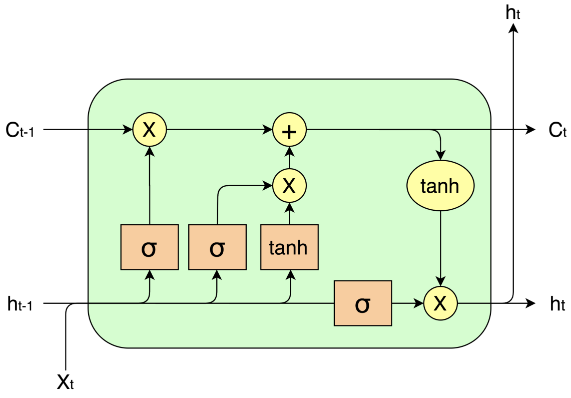
Anonymous ACL submission

1 Description of the dataset

The dataset comprises two tasks aimed at detecting hate speech. Task 1 focuses on Hate Speech Detection, where the goal is to classify whether a message contains hate speech or not. The training set consists of 5,600 tweets from PolicyCorpusXL, while the test set includes 1,400 tweets from PolicyCorpusXL and 3,000 tweets from ReligiousHate. In Task 2, termed Contextual Hate Speech Detection, both the content of tweets and their metadata are considered for classification. This task includes two sub-tasks: Political Hate Speech Detection, which utilizes data from both development and test sets from PolicyCorpusXL, and Religious Hate Speech Detection, where only the test data from ReligiousHate is provided, adapted from an original cross-domain task.

2 Architecture of your model

My model is a bidirectional LSTM with a projection layer at the end.



3 Design choices of your model

I used a low number of layers and the hidden size because I noticed that the dataset was quite easy to calssify so it didn't need a too complex model

4 Baselines implemented

I implemented 2 baselines, a random one, that simply generates a random tensor of the desired dimensions and an RNN

5 Results section

LSTM:

$$Epoch : 20$$

train loss = 0.6384

$$Epoch : 20$$

valid loss = 0.6380, valid acc = 0.9961 Test loss 0.6359703506742205, Test accuracy: 0.9959821428571428

RNN:

$$Epoch : 20$$

train loss = 0.5663

$$Epoch : 20$$

valid loss = 0.5644, valid acc = 0.8583

Test loss 0.5634483064923967, Test accuracy: 0.8554315481867109

I noticed that the results vary drastically depending on the initialization, it can either overfit from the beginning or starting from a bad initialization might mean that the initial accuracy is 0.000, and that's for both the LSTM and RNN.

6 Instructions to run your code

To run the training you should set the folder contating hw1b_train.py as CWD and then run python hw1b_train.py as for the evaluation you should get to the same folder and run python hw1b_evaluate.py