

Report1

The model's architecture is as follows:

Every char embedded to a 10-dimension vector, which then was input as a sequence to the lstm. The lstm output is a 100-dimension vector. This vector was used as an input to the MLP, which has one hidden layer of size 50, and its output activation is sigmoid which outputs a scalar y_{pred} on the interval $[0,1]$. If $y_{pred} \leq 0.5$ then the model's prediction is negative example, else positive example.

The training size is 800, validation size is 100 and test size 100.

The total epochs we used is 100 and also used early stopping on the validation loss.

Total training time is 13.3 seconds with batch sizes of 32.

The model was able to distinguish (sometimes, depends on the randomness) with perfect accuracy on the test set.