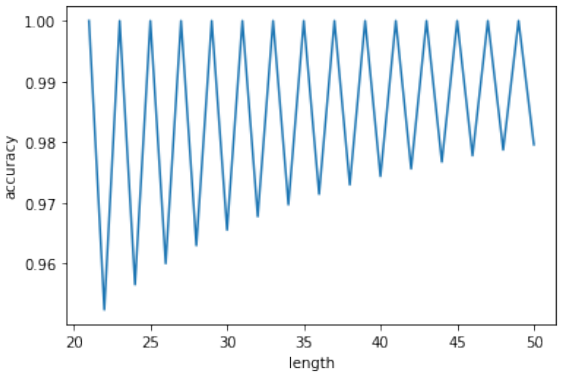
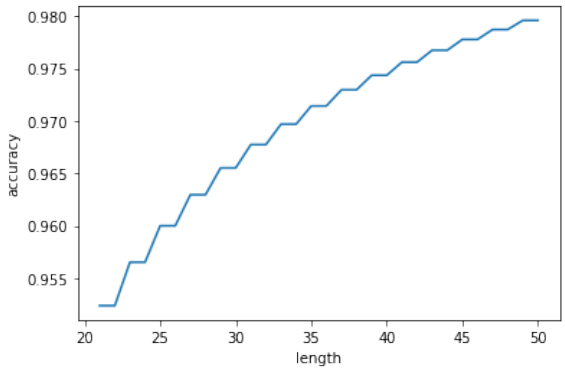
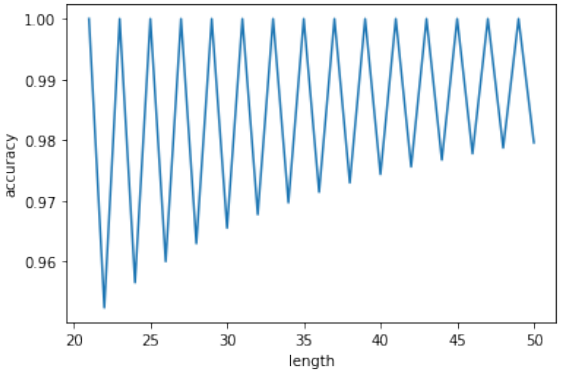


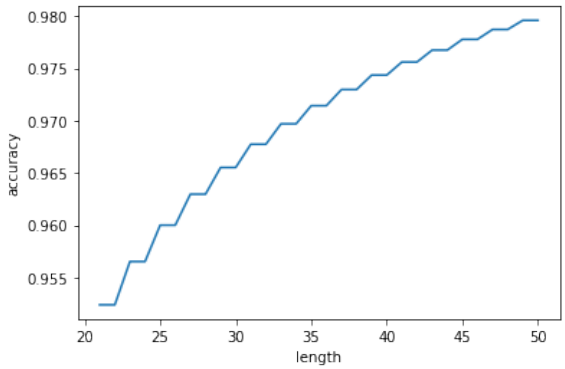
(a) lstm-ab



(b) lstm-abc



(c) rnn-ab



(d) rnn-abc

Hyper-parameter search by splitting the training data:

Both of these two models randomly selects 10% of the data-set to be validation data and keeps the rest of data-set as the training set. I decide the hyper-parameters by searching for materials online and also my own trial and error. I found the the split of training and validation set does not influence the accuracy of validation test – they are always 100% accurate. For other parameters like batch size(10000) and buffer(1000) size, I just use the one that can cover all the vocabulary. Other settings just follow the default one since the overall performance does not increase a lot when I change them, e.g Adam(1e-4). The epochs are 100 since we can still see the increase of accuracy for smaller ones.

Evaluate models on longer languages:

According to the graphs, we can see that the overall accuracy of all models are quite good: they are all over 95% and the performance increases with the increase of language length.Even if there is small fluctuation, the overall performance converge to a higher level, interestingly, there also exits a difference between odd length and even length. The accuracies of RNN and LSTM are similar, though lstm is slightly better.