Report on Sentence-level Sentiment Classification with RNN

千英卓 2017013622

Introduction

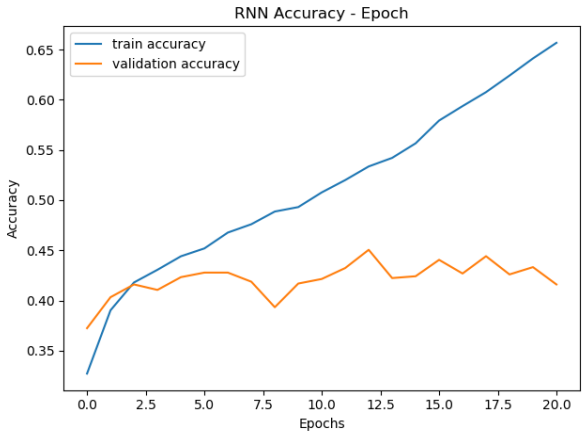
This experiment aims to perform sentiment classfication on sentences using RNN and its variables. We're expected to complete the framework of various RNN network provided by TA. The implementation of self-attention is also required.

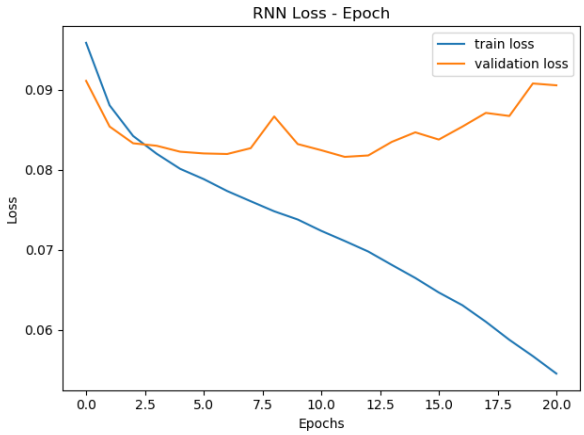
Experiment Result

One-layer Models

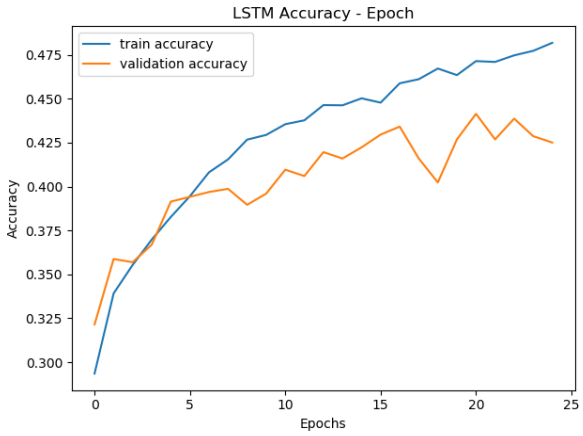
The figures of loss and accuracy during training and performance of one-layer RNN models is shown below.

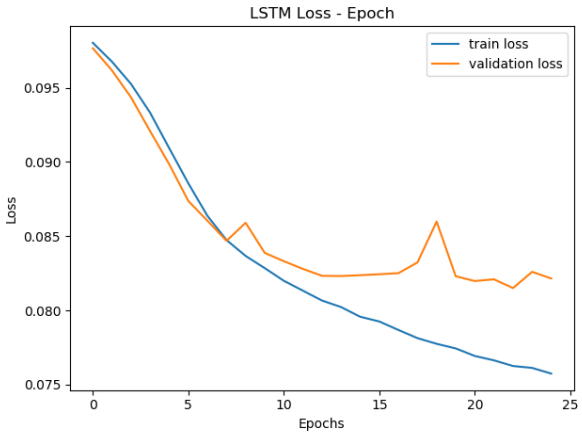
Basic RNN:



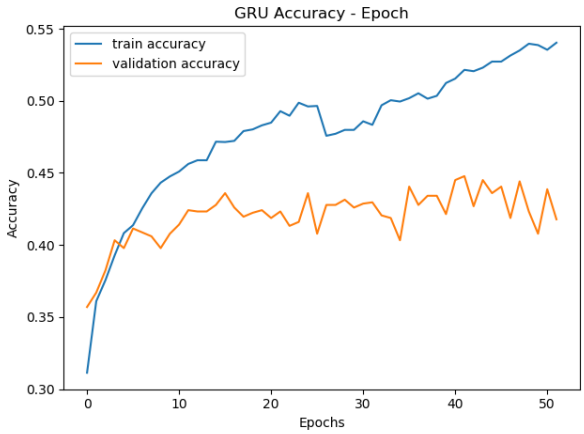


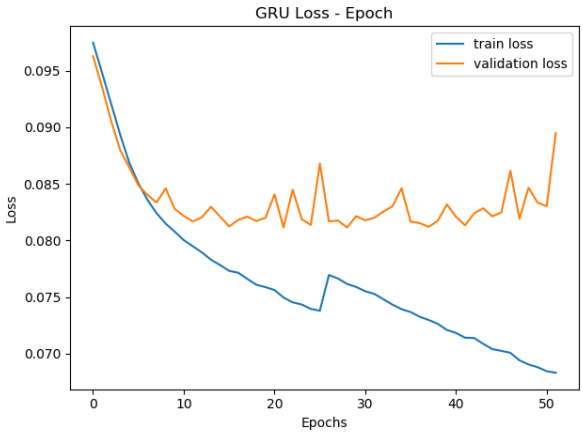
LSTM:





GRU:





Performances

Model	Highest Validation Accuracy	Training Time(s)
Basic RNN	0.4505	2754
LSTM	0.4414	7322
GRU	0.4478	10546

Note: the training stops either after 50 epochs or after a consecutive 8 epochs without reaching new highest validation accuracy.

It can be seen that among the three, basic RNN model has the highest validation accuracy. It also uses the least amount of time to train, which is expected as the calculation inside one cell is the least complicated. However, the difference between 3 models is somewhat modest, therefore the result may be arbitrary to some extends, considering that the size of dev set is rather small.

Also, the huge margin between accuracy of train set and validation set indicates that over-fitting is severe, probably due to the skewed distribution.

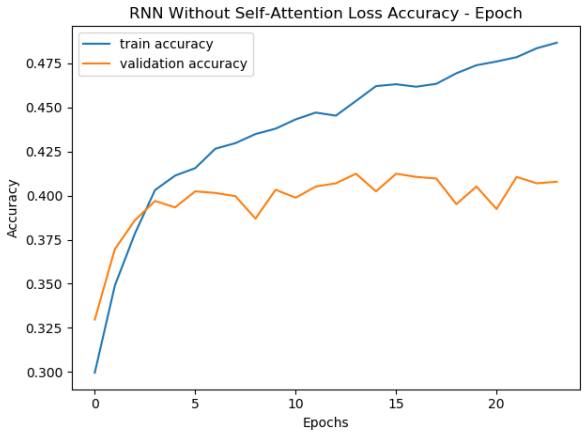
Models without Self-Attention

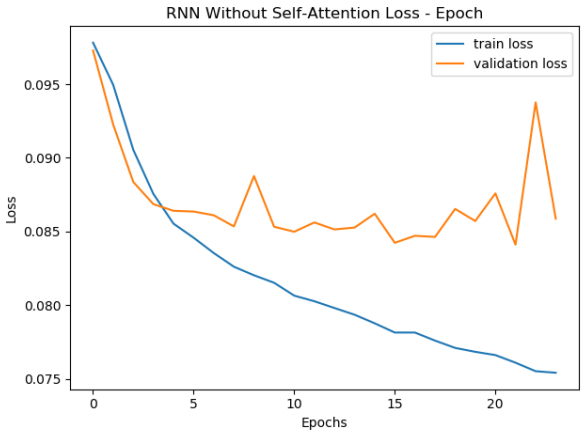
Instead of self-attention technique, a average pooling layer is placed after the output tensors of RNN:

```
M = tf.reduce_mean(H, axis=1)
```

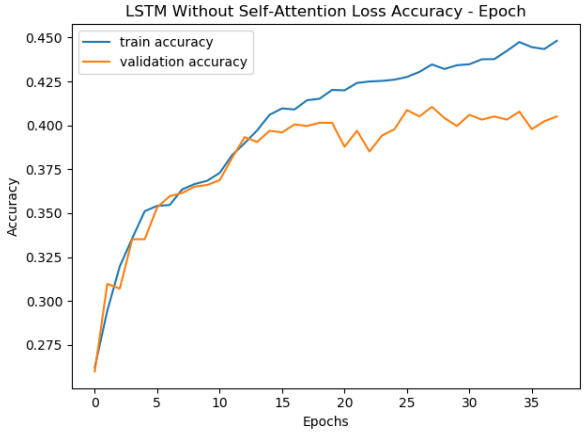
The figures of loss and accuracy during training and performance of one-layer RNN models is shown below.

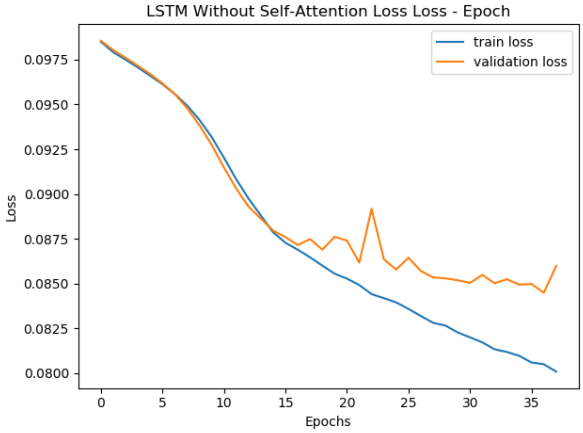
Basic RNN:



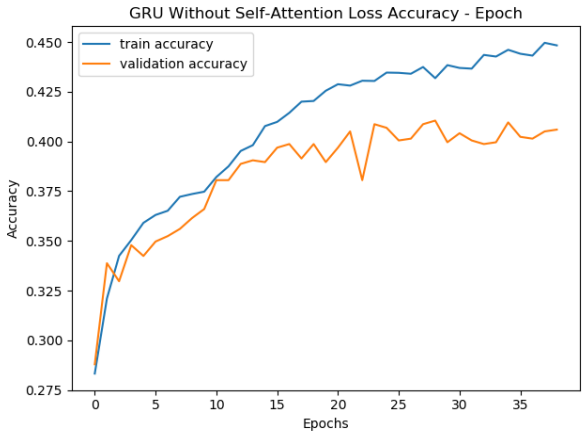


LSTM:





GRU:



Performances:

Model	Highest Validation Accuracy	Training Time(s)
Basic RNN	0.4123	1529
LSTM	0.4105	10013
GRU	0.4105	8018

Note: the training stops either after 50 epochs or after a consecutive 10 epochs without reaching new highest validation accuracy.

It can be seen that the performance of models without self-attention is significantly inferior to that of those that have, which is expected. The highest validation accuracy of all 3 models is considerably lower though more training time is required.

Final Model

Though one-layer basic RNN has the highest validation accuracy among all models, it seems that basic RNN tends to overfit very quickly. The train accuracy of the model promptly goes above 0.65 within 20 epochs of training, while the train set accuracy of GRU and LSTM is under 0.50. Thus, I choose one-layer GRU to be my final model.

The loss and accuracy of train and validation set is ploted above. As can be seen, the model inevitably has serious overfitting problem, but not as sever as that in basic RNN model. It reaches second highest validation accuracy though requires longer time to train.