IST 597: Assignment 10000

Jooyoung Lee
Information Sciences and Technology
Pennsylvania University
920172979

1 Problem 1:Implementing various Recurrent neural network cells using basic tensorflow ops

1. Gated Recurrent Unit (GRU)

Update equation for Gated Recurrent Unit(GRU). GRU is the simplified version of LSTM and is also as powerful as LSTM in terms of generalization performance. We encourage you to read [1] for in-depth understanding.

rt =(Wr[ht1,xt]+br) s =Tanh(W[rs ,x]+b) t s t t1 t s s =(1z)s +z s t t t1 t t

$$\begin{aligned} \mathrm{z}_t &= \sigma(W_z[h_{t-1}, x_t] + b_z) \\ \mathrm{r}_t &= \sigma(W_r[h_{t-1}, x_t] + b_r) \\ \hat{s}_t &= \tanh(W_s[r_t \odot S_{t-1}, x_t] + b_s) \\ \mathrm{s}_t &= (1 - z_t) \odot s_{t-1} + z_t \odot \hat{s}_t \end{aligned}$$

2. Minimal Gated Unit (MGU)

Below we have provided the equations for minimal gated unit(MGU). MGU is the simplified version of gated recurrent unit. MGU is 3 times faster than LSTM and 2 times faster than GRU due to less update calls.We encourage you to read [2] for in-depth understanding.

$$\begin{aligned} \mathrm{f}_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ \hat{s}_t &= \tanh(W_s[f_t \odot S_{t-1}, x_t] + b_s) \\ \mathrm{s}_t &= (1 - f_t) \odot s_{t-1} + f_t \odot \hat{s}_t \end{aligned}$$

1.1 Result

In order to compare the performance of BasicLSTM, GRU and MGU on NotMnist dataset, I retrieved the loss and accuracy of these models. The total number of dataset was 529114 images that consist of 10 classes and I used 20 percent of the whole dataset as my test set. In terms of hyperparameters, I set minibatch size as 100, learning rate as 0.01 and hidden units as 128. Also, I implemented AdamOptimizer for the optimization function.

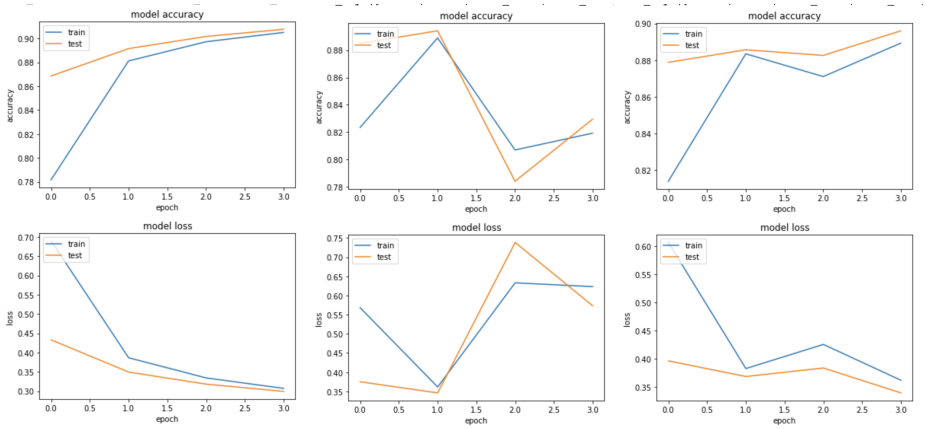
loss	BasicLSTM	GRU	MGU
epoch1	0.6799	0.6067	0.5986
epoch2	0.3735	0.3831	1.4616
epoch3	0.3363	0.4259	1.0202
epoch4	0.3003	0.3624	1.6264
Test Loss	0.29688038	0.5733146	0.3399031

[Fig1] loss report using 3 different models

acc	BasicLSTM	GRU	MGU
epoch1	0.7828	0.8139	0.8164
epoch2	0.8856	0.8835	0.5711
epoch3	0.8958	0.8710	0.6894
epoch4	0.9066	0.8893	0.5263
Test Acc	0.9092	0.82945	0.89598

[Fig2] accuracy report using 3 different models

The results show that loss is minimum when using a LSTM model compared to GRU and MGU. Also, the accuracy is maximum when using a LSTM model as well. Even though loss seems to go up and accuracy seems to go down after epoch2 for MGU, its final accuracy achieved a robust performance compared to GRU. Moreover, the computational time of MGU was the fastest since it is simpler than other two models.



[Fig3] learning curve of LSTM, GRU, MGU from left to right

Figure 3 displays the learning curve of three different models and GRU's curve appears to be fickle and inconsistent whereas a BasicLSTM model and a MGU model relatively shows a constant decrease/increase in both loss and accuracy.