EE599 Deep Learning – Homework 5

1 How to run the python files:

Run the data.py:

to create the train set /test set with MFCC features, which is named: ***mfcc_dataset.hdf5***.

(about 7.5G)

Set the configuration in utils2.py:

Input_path: path of the file "train"
Num_mfcc_features: the number of the mfcc features
Tetsset_size: the ratio of test set and train set
Debug: because the test.py will take a lot of time so if debug == true, only make one prediction.
Model: choose the model used in the train.py: CuDNNGRU/CuDNNLSTM

Run the train.py:

Get the model based on my training: named: ***model_LSTM.hdf5/ model_GRU.hdf5***

Run the test.py:

If debug is True only make one prediction and show both the true label and the predict label.
If debug is False, make all prediction in the test set and count the number of each three languages predicted.

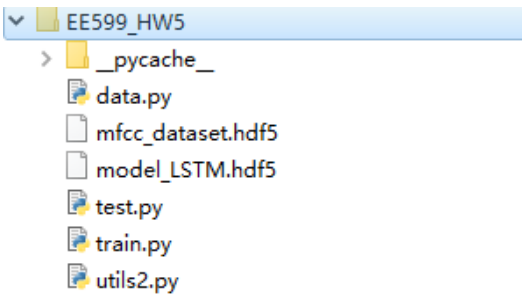

Figure 1

2 How to get the final model:

Firstly, I started with the file: toy_rnn.py and do some improvement that can be used in this task, but I found it's very slow in training and the accuracy is around 0.6 after about 75 epochs. Then I found this package: tf.compat.v1.keras.layers.CuDNNGRU which can only be run on GPU and it is much faster than the GRU.

I used only GRU layer and one Dense layer but found the validation accuracy is better than before but not good enough around 0.7 finally. So I add a GRU layer and a Dense layer, which is the final model (showed in figure 2), and it achieve **0.8897** validation accuracy finally, and **0.9608** training accuracy finally. (showed in figure 3)

At last the way that I handle silence is the first way just label the silence as the languages because I think that the length of silence of different language is different and it may be a very good feature to do the classification.

Layer (type)	Output Shape	Param #
main_input (InputLayer)	[(None, 1000, 64)]	0
layer1 (CuDNNGRU)	(None, 1000, 64)	24960
layer2 (CuDNNGRU)	(None, 1000, 32)	9408
layer3 (Dense)	(None, 1000, 100)	3300
rnn_output (Dense)	(None, 1000, 3)	303

Total params: 37,971
Trainable params: 37,971
Non-trainable params: 0

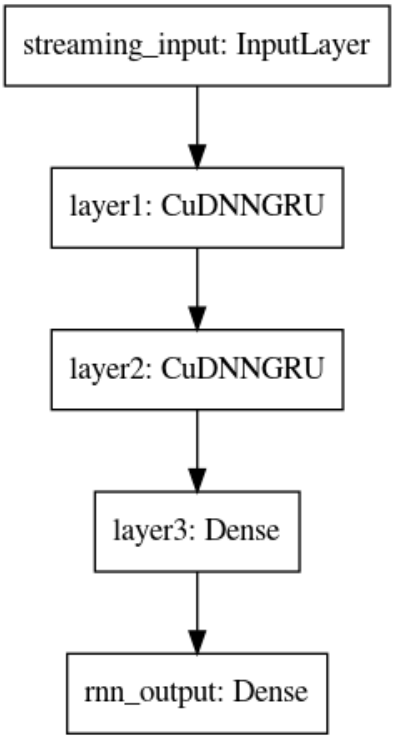

Figure 2

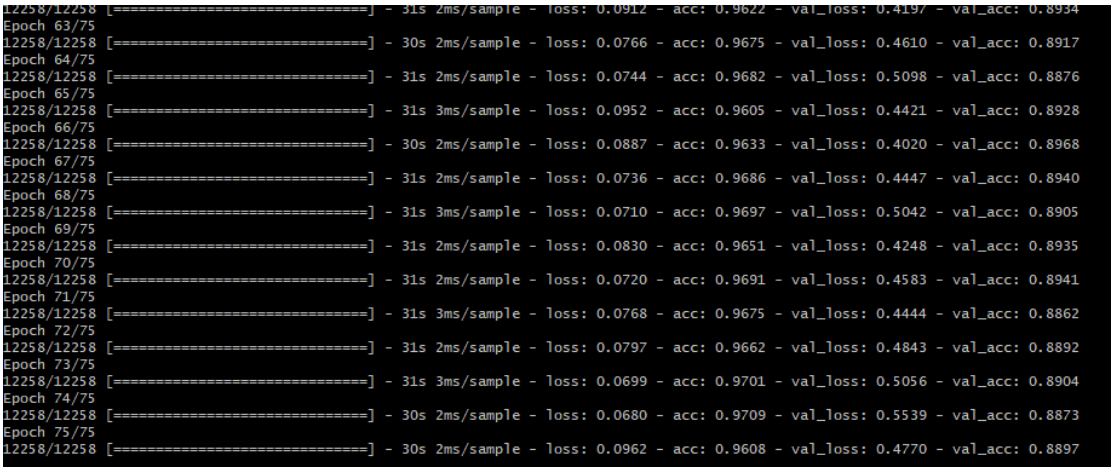
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0912 - acc: 0.9622 - val_loss: 0.4197 - val_acc: 0.8934
Epoch 63/75
12258/12258 [==============================] - 30s 2ms/sample - loss: 0.0766 - acc: 0.9675 - val_loss: 0.4610 - val_acc: 0.8917
Epoch 64/75
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0744 - acc: 0.9682 - val_loss: 0.5098 - val_acc: 0.8876
Epoch 65/75
12258/12258 [==============================] - 31s 3ms/sample - loss: 0.0952 - acc: 0.9605 - val_loss: 0.4421 - val_acc: 0.8928
Epoch 66/75
12258/12258 [==============================] - 30s 2ms/sample - loss: 0.0887 - acc: 0.9633 - val_loss: 0.4020 - val_acc: 0.8968
Epoch 67/75
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0736 - acc: 0.9686 - val_loss: 0.4447 - val_acc: 0.8940
Epoch 68/75
12258/12258 [==============================] - 31s 3ms/sample - loss: 0.0710 - acc: 0.9697 - val_loss: 0.5042 - val_acc: 0.8905
Epoch 69/75
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0830 - acc: 0.9651 - val_loss: 0.4248 - val_acc: 0.8935
Epoch 70/75
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0720 - acc: 0.9691 - val_loss: 0.4583 - val_acc: 0.8941
Epoch 71/75
12258/12258 [==============================] - 31s 3ms/sample - loss: 0.0768 - acc: 0.9675 - val_loss: 0.4444 - val_acc: 0.8862
Epoch 72/75
12258/12258 [==============================] - 31s 2ms/sample - loss: 0.0797 - acc: 0.9662 - val_loss: 0.4843 - val_acc: 0.8892
Epoch 73/75
12258/12258 [==============================] - 31s 3ms/sample - loss: 0.0699 - acc: 0.9701 - val_loss: 0.5056 - val_acc: 0.8904
Epoch 74/75
12258/12258 [==============================] - 30s 2ms/sample - loss: 0.0680 - acc: 0.9709 - val_loss: 0.5539 - val_acc: 0.8873
Epoch 75/75
12258/12258 [==============================] - 30s 2ms/sample - loss: 0.0962 - acc: 0.9608 - val_loss: 0.4770 - val_acc: 0.8897

Figure 3