

REPORT DEEP LEARNING & TIME SERIES

Nội dung:

- Nhận diện số 1-9 sử dụng deep learning với time series
- Ngôn ngữ sử dụng: Python 3.6

I. Phương pháp:

1. Feature extraction: Sử dụng MFCCs
2. Training: Sử dụng một mạng Recurrent neural network(RNN). Ở đây dùng mạng LSTM để thực hiện training:
3. Test

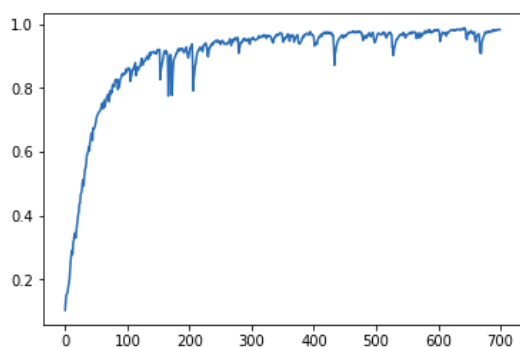
II. Chi tiết:

1. Mô tả data:
 - mỗi số có khoảng 360 sample cho cả giọng nam và giọng nữ
 - file .wav $F_s = 16000$.
 - Mỗi file có độ dài khoảng 2.5s
2. Feature extraction:
 - Tách file .wav ra thành nhiều frame. Mỗi frame 0.025s
 - Mỗi frame lấy ra 26 MFCC coefficients
 - Stack lại thành vector dataset
3. Training:
Cấu trúc model:

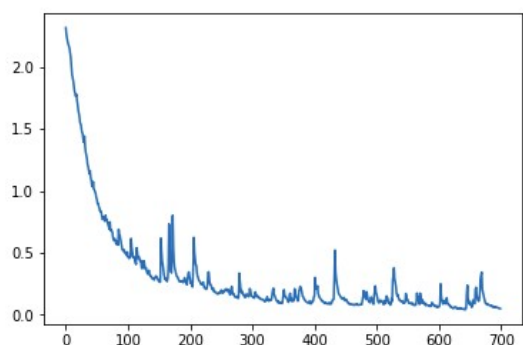
Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 26)	5512
dropout_1 (Dropout)	(None, 26)	0
dense_1 (Dense)	(None, 128)	3456
dropout_2 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 10)	1290
activation_1 (Activation)	(None, 10)	0
Total params: 10,258		
Trainable params: 10,258		
Non-trainable params: 0		

- Kết quả sau khi train 700 epochs độ chính xác 98%:

Accuracy



Loss



III. Link tham khảo:

Dataset:

https://storage.cloud.google.com/download.tensorflow.org/data/speech_commands_v0.02.tar.gz

MFCCs: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>

LSTM: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>