NLP Application Project

Urvashi Senha
201225020

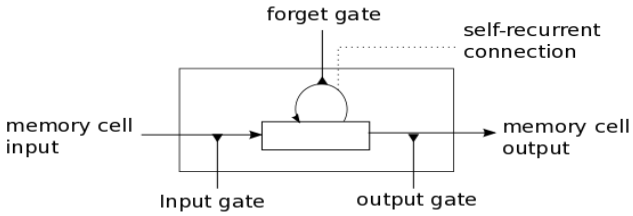
Title : Grapheme_to_Phoneme Aligner using LSTM Network

Introduction :

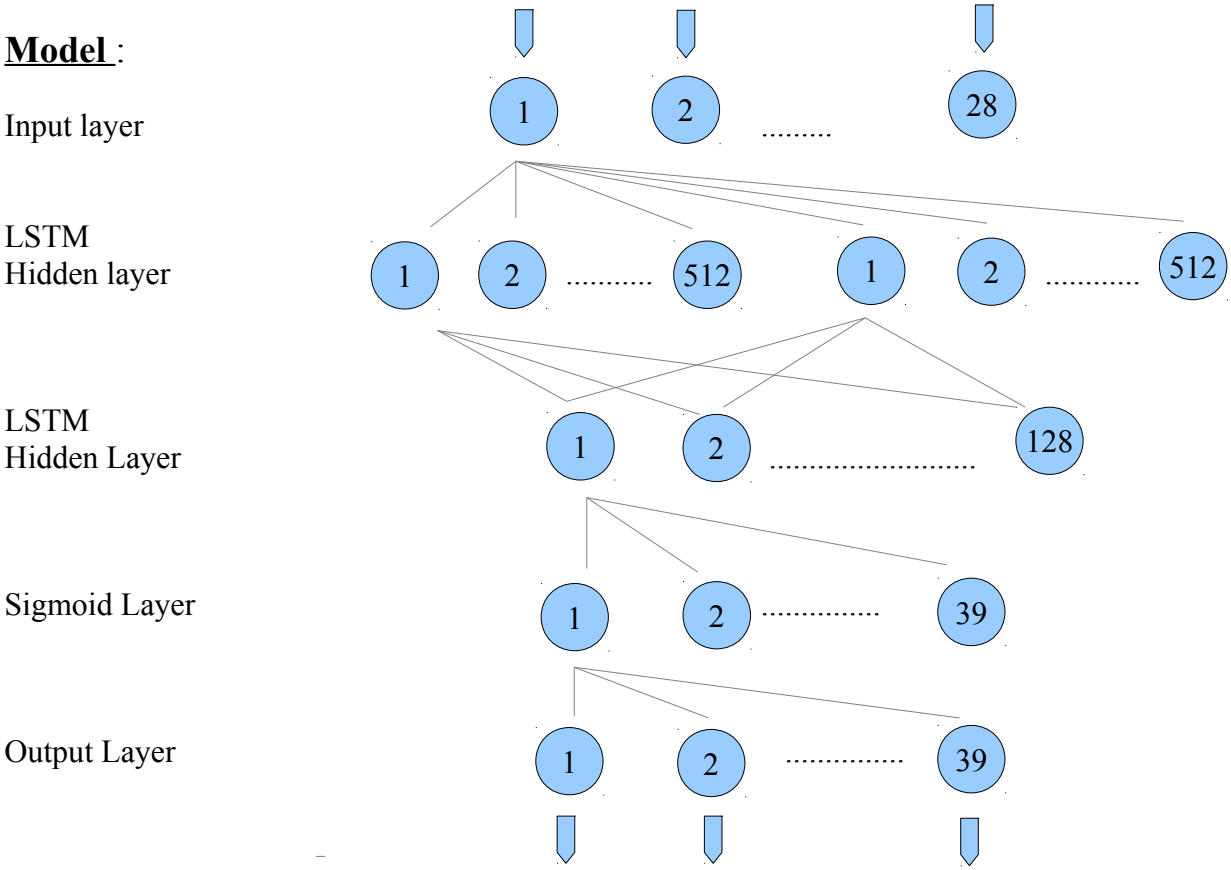
Grapheme-to-phoneme (G2P) models are key components in speech recognition and text-to-speech systems as they describe how words are pronounced. A G2P converts a word, as a series of characters or graphemes, to a prounciation, as a series of phones . For the project, I have implemented a Grapheme-to-phoneme model using LSTM RNN approach.
For data, CMUDict is used. The input grapheme set is the 26 English Alphabet and 2 symbols : apostrophe (') and hypen (-) . The output phoneme set is ARPA (39 phonemes).

LSTM Neuron :

LSTMs avoids the need for explicit alignment before training; instead, with a dynamic contextual window, the LSTM may see several graphemes before outputting any phoneme, which allows it to make contextually- aware decisions.



Model :



INPUT - The input vector is a 28 dimension vector, “one-hot”representation of the input grapheme sequence.

OUTPUT – The output vector is a 39 dimensional vector, “one-hot” representation of the output phoneme sequence.

Hidden Layer 1 – 2 parallel layers each if 512 neurons
Hidden Layer 2 – consisting of 128 neurons
Sigmoid Layer – consisting of 39 neurons

All connections between the layers are full connections.

Results:

Training Samples	Epochs	Learning Rate	Test Error Rate	Training Error Rate
500	100	0.01	8%	0.3%
		0.02	8%	0.2%
500	500	0.01	8%	0.13%
		0.02	8%	0.12%
1000	500	0.02	4%	0.1%

Conclusion:

- The hidden units if reduced were causing an increase in the training error rate.
- The learning rate of 0.02 was found to be the most apt.
- The minimum Test Error rate was 4% with 500 epochs, and 1000 training samples. Increasing the Epochs lead to no significant improvement.

Future Work :
Extracting the optimal sequence using FST.

References :
GRAPHEME-TO-PHONEME CONVERSION USING LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORKS - Kanishka Rao, Fuchun Peng, Hasim Sak, Francoise Beaufays