Final Project Proposal Part 2

1. For my final project I will utilize an LSTM recurrent neural network. Neural networks are described in chapter 7 of our textbook. The input feature vector will comprise of mel frequency cepstral coefficients of a wav file, representing a certain vowel phoneme. The output layer will comprise of a vector that defines a certain vowel phoneme prediction.
2. My network will be evaluated with training loss metrics that are offered by Keras’ LSTM library. Specifically, I intend to use a method labeled ‘categorical\_crossentropy’ to measure the loss of my algorithm. This method takes in true labels and predicted labels and then calculates the cross-entropy value for my multiclass classification problem.
3. The following link provides an implementation of an LSTM network with Keras for python, for performing sentiment analysis: <https://towardsdatascience.com/machine-learning-recurrent-neural-networks-and-long-short-term-memory-lstm-python-keras-example-86001ceaaebc>. From this webpage I navigated to the Keras homepage and found three useful links for the implementation of this specific package:
   1. <https://keras.io/models/sequential/>
   2. <https://keras.io/metrics/>
   3. <https://keras.io/layers/recurrent/#lstm>
4. The process I am interested in has been researched in the following papers:
   1. <https://arxiv.org/pdf/1810.04635v1.pdf>
      1. This paper uses multiple models for speech emotion recognition including a convolutional LSTM network.
   2. <https://arxiv.org/pdf/1806.01506v2.pdf>
      1. This paper is similar to the latter, regarding speech emotion recognition, however they utilize a spectrogram representation of audio files for the input of a fully convolutional network.
   3. <https://link.springer.com/chapter/10.1007/11550907_126>
      1. This paper falls the most in line with my project as it uses a bidirectional LSTM network for phoneme classification.

Citations

1. Graves, Alex, et al. “Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition.” *SpringerLink*, Springer, Berlin, Heidelberg, 11 Sept. 2005, link.springer.com/chapter/10.1007/11550907\_126.
2. “Keras: The Python Deep Learning Library.” *Home - Keras Documentation*, keras.io/.
3. Yoon, Seunghyun, et al. “Multimodal Speech Emotion Recognition Using Audio and Text.” *2018 IEEE Spoken Language Technology Workshop (SLT)*, 2018, doi:10.1109/slt.2018.8639583.
4. Zhag, Yuanyuan, et al. “Attention Based Fully Convolutional Network for Speech Emotion Recognition.” *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2018, doi:10.23919/apsipa.2018.8659587.
5. Maklin, Cory. “LSTM Recurrent Neural Network Keras Example.” *Medium*, Towards Data Science, 21 July 2019, towardsdatascience.com/machine-learning-recurrent-neural-networks-and-long-short-term-memory-lstm-python-keras-example-86001ceaaebc.