Tristin Johnson

DATS 6450 – Machine Learning II

December 6th, 2021

Final Project Proposal

- **What problem did you select and why did you select it?**

  The problem that I selected for my final project is a problem set called Urban Sounds. The topic of this problem is under Speech Recognition, specifically Deep Speech, in which the goal of this project is to correctly classify audio files using a neural network. I selected this project because I wanted to work with audio files and implement a Deep Speech neural network as it is a topic that has always interested me. I am also curious in how to transfer the audio files into a neural network and what the workflow of these steps include.

- **What database/dataset will you use? Is it large enough to train a deep network?**

  The dataset that I will use is called UrbanSounds8K. This dataset consists of just over 8000 audio files (in .wav format) and the data includes 10 different classes (list all the classes here). Each audio file is anywhere between 3.75 seconds to 4 seconds long. Therefore, with over 8000 audio files, this dataset is most definitely large enough to train a deep network.

- **What deep network will you use? Will it be a standard form of the network, or will you have to customize it?**

With the research I have done thus far, a CNN seems to be the best fitted model for Deep Speech. This network will most definitely have to be customized, as I believe the standard form of the network for this dataset is quite simple with low levels of accuracy. There also is room to try and implement an LSTM neural network. Furthermore, after implementing a CNN model, I would like to test the network using a few different pre-trained models that exist within Deep Speech and compare that performance to the performance of my model.

- **What framework will you use to implement the network? Why?**

  The framework that I plan on using to implement the network is PyTorch. In the past, I have used Tensorflow for most of my projects related to neural networks. Recently, I was introduced to PyTorch and enjoy the implementation of PyTorch because I feel as if I have more control on the specifications of the model. Furthermore, after doing some research on Deep Speech, a lot of different research papers mention that PyTorch was used as the framework as PyTorch includes multiple torch-specific packages for working with audio files (such as torchaudio and torchvision).

- **What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?**

  I plan on doing a lot of research and looking at several different Deep Speech models using various references on the internet. I plan on following a few different tutorials in working with audio files (such as PyTorch audio datasets) to get comfortable in working

with .wav files, converting the audio to Mel Spectrograms, data augmentation, and a few other techniques.

- **How will you judge the performance of the network? What metrics will you use?**

  To judge the performance of this network, the accuracy score will be the best metric for this specific problem. Looking at the data, there are 10 different classes and 8 of those classes have the exact same number of audio files (1000 files each), which means there is an even distribution across 8 of the classes. The other 2 classes do have less than 1000 files, but not too far behind. Therefore, since there is very little class imbalance among the dataset, accuracy will be the best judgement, and F-1 Score may also be applied.

- **Provide a rough schedule for completing the project.**

  The first step is to get comfortable with the dataset, see what the audio files look like, and figure out the best way to convert .wav files into valid inputs for a neural network. This should be done within the next two weeks (by November 8th, 2021). The next step is to create and implement a working CNN model that has the ability to train and test the dataset, along with identifying the audio files to their correct class, and this should be done before Thanksgiving break (November 22nd, 2021). Once there is a working model, the next 2-3 weeks will be spent tweaking and modifying the model to get the highest levels of accuracy possible (such as data augmentation, different CNN architectures, applying different optimizers, etc.), which will be complete by the due date of the final project, December 6th, 2021.