Matthew Perez

CMPLXSYS

Assignment 2

1)

I’m very interested in modeling human health. Humans are such complex creatures and I’ve always been fascinated with how technology can be used help understand what’s happening within the body. Complex systems are relevant to this topic because the human system is complicated on both an internal and external level. Externally, there are a variety of different input and output signals to observe. For example, Facial images can study appearance while EKG can provide a much lower and more fundamental signal. Choosing which signal to observe takes a lot of forethought and ultimately depends on the end goal (How will this be used in practice? which is easier to collect? Etc.). Additionally there are many different methods, which can be used to model the human system. As we’ve previously discussed in class, an agent-based model might be useful for simulating how blood responds to a wound or infection. However, as a computer scientist I am particularly interested in using neural networks as my model of choice because of their recent success and power to map higher order functions. My end goal is see how I can model human output with respect to neurological disease. Specifically, my goal is to study how audio extracted speech cues can be used to characterize Huntington Disease.

2)

Previous speech research by Fraser et. al. has shown that for Aphasic speech, observing the content of what was said through transcript extracted features achieves high classification performance[1,2]. Although, these transcript features are promising, they rely on someone to manually transcribe what was being said. Work here at the University of Michigan by Le et. al. investigated the use of automatic speech recognition (ASR) for generating these transcripts and ultimately performing aphasic classification.[3] It is important to train an ASR system that is as accurate as possible since transcript accuracy will lead to higher quality features.

Similar to Aphasia, HD speech symptomatology involves slower speaking rate, longer and more frequent pauses in speech, and poor pronunciation. My previous research has investigated how these specific features can be extracted from ASR generated transcripts and used to identify the presence of HD.[4] However, HD also has been shown to affect respiratory and articulatory functions resulting in nasaly and jittery speech.[5] Traditional approaches for capturing this symptomatology involve extracting low-level acoustic features directly from the speech signal itself.[6]

3)

I want to focus on Huntington Disease because my lab has access to a private dataset collected here at the University of Michigan Medical School, which consists of audio and transcripts from speakers with Huntington Disease. This dataset has 62 speakers, 31 healthy and 31 with HD. I’m interested in this dataset because the data collected was based off of participants reading the Grandfather Passage.

I plan to extract both transcript-generated features (speech rate, pronunciation, pauses) and compare the effects of these features to low-level acoustic features, which will capture rhythm, jitter, and shimmer. Additionally can these features be improved by studying segmented portions? How do the effectiveness of these features change with the added knowledge of demographic information? The model as mentioned earlier will be neural networks where the agents can be thought of as neurons in the network. These neurons will attempt to capture important relationships between the identified input features and class outputs. I’m not sure what the environment is in the case of neural networks and what role they play. Agents don’t move through the environment but do interact with a specific set of other agents.

I have done similar research in python, which extract acoustic features and train a system for automatic speech recognition (another model). I also have experience in extracting speech-based features using the transcript and audio file (forced alignment or speech recognition), which I can then use to classify the disease. For the purpose of this project however, I will be investigating new features, new approaches, and performing feature analysis to investigate which features are important to understanding HD and can these features be used to classify HD severity.

References:

[1] Fraser, Kathleen C., Frank Rudzicz, and Elizabeth Rochon. "Using text and acoustic features to diagnose progressive aphasia and its subtypes." *INTERSPEECH*. 2013.

[2]Fraser, Kathleen, et al. "Automatic speech recognition in the diagnosis of primary progressive aphasia." *Proceedings of the fourth workshop on speech and language processing for assistive technologies*. 2013.

[3] Duc Le and Emily Mower Provost. “Emotion Recognition From Spontaneous Speech Using Hid- den Markov Models With Deep Belief Networks.” Automatic Speech Recognition and Understanding (ASRU). Olomouc, Czech Republic. December, 2013.

[4] Perez, Matthew, et al. "Classification of Huntington Disease using Acoustic and Lexical Features." *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*. Vol. 2018. 2018.

[5] Hamilton, Alison, et al. "Management of speech, language and communication difficulties in Huntington’s disease." *Neurodegenerative Disease Management* 2.1 (2012): 67-77.

[6] Syed, Zafi Sherhan, Kirill Sidorov, and David Marshall. "Automated Screening for Bipolar Disorder from Audio/Visual Modalities." *Proceedings of the 2018 on Audio/Visual Emotion Challenge and Workshop*. ACM, 2018.