Machine Learning Qualifying Exam Report *Georgia Institute of Technology*

PhD Student
Nicolas Shu

Committee Members:
David Anderson
Gari Clifford (Advisor)
Evangelos Theodorou

Contents

1 Background

Automatic speech recognition (ASR) has been a topic that interested many from an early age. Many consider it to have started in the 1950s with Bell Labs' *Audrey* [Moskvitch, 2017], which was able to do a single-speaker digit recognition. Seventy years later, the ASR technologies have grown so that they are present within our personal homes, with Amazon Alexa and Google Home devices [Fowler, 2018], as they have become more affordable. As the speech technologies advance, they approach the concept of ubiquitous computing (often also known as ambient intelligence), which is highly desirable for many industries. Although most of the ASR systems available for commerce are trained in normal language and not medical language, one industry that could greatly benefit from ASR technologies is the health care system, given that the ASR system is able to understand medical language.

Health care is a system which has a very high demand and the services are required to be as detailed as possible due to various reasons, one of which is the concept that it is paramount for a hospital to have a clear and rich track of a patient's medical history. Currently, in order to maintain a patient's medical history, a physician sees multiple people during his/her working hours, and only after their shift is over, does (s)he sit down to write the medical notes. The lag in between seeing a patient and taking notes may sometimes go up to 8-10 hours, and then those notes are often inaccurate. Having an ASR system in a medical environment could greatly help in keeping track of a patient's medical history, where a physician could easily dictate the notes. In a hospital environment, however, such as in an intensive care unit (ICU), a physician who is trying to dictate notes may find him or herself in trouble, as in the ICU, there are multiple background sounds from machines, multiple people speaking, and a great amount of white noise.

Current companies have been creating ASR systems that can understand medical language and allows physicians to dictate their medical notes. One major company that has been "dominating" a lot of the market is Nuance, with their Dragon Medical system. Unfortunately, their dictation system is not yet capable of inferring punctuation marks and markup language onto the text, thus, in order to dictate a segment such as: "37-year-old female presents complainint of urinary frequency, urgency and dysuria along with hematuria and low-grade fever." needs to be dictated as:

"37-year-old female presents complainint of urinary frequency **comma** urgency and dysuria along with hematuria and low-grade fever **period**"

Although the system has very high accuracy results, the system works best when one is in a quiet environment, which is not always a realistic scenario. The goal of this project is to create a transcription engine that can recognize medical language in a noisy environment.

2 Preliminary Data

2.1 Comparison of Different Commercial Engines

Compare commercial (+ CMU) Text-¿Voice-¿ Text MIMIC II + 20kLeagues

2.2 Grade Reading Level of Various Texts

MIMIC II + 20kLeagues

3 Automatic Speech Recognition Schematic for a Medical Setting

In order to have a speech recognition system to be able to understand medical language, it is important to obtain a big picture of the project.

[INSERT FIGURE]

As one can see from the diagram above, the sound signals would come to a microphone or an array of microphones, which would then be passed through a system that performs blind source separation on the sound signal. It is very possible that, in a hospital environment, there may be multiple sources, such as the heart rate monitor, the ventilator, the healthcare providers talking, and others. A blind source separation algorithm would ensure to separate the audio channel to multiple sources. Once the sources have been separated, there would be a classification methodology to identify the physician from the patient, and ...

Although there have been many advancements in the speech technologies and there currently exists very accurate engines to do voice recognition, it still remains to be an unsolved problem, as there is always room for improvement. Therefore, there are a lot of possibilities for the formulation design of the framework (i.e. pipeline).

4 Speech Enhancement

4.1 Convolutional Neural Networks for Noise Suppression

4.2 Non-Negative Matrix Factorization

Please skip to the Blind Source Separation's subsection on Non-Negative Matrix Factorization

5 Blind Source Separation

5.1 Independent Component Analysis

5.2 Non-Negative Matrix Factorization (NMF) and Sparse NMF

6 Speaker Diarisation

7 Speaker Verification

8 Speech Recognition

8.1 Hidden-Markov Models and Gaussian Mixture Models

8.2 Long Short Term Memory Networks

[Shwartz-Ziv and Tishby, 2017] [Chaudhari et al., 2018] [E et al., 2018]

9 Deep Relaxation: PDEs for Optimizing Deep Neural Networks

10 Characterization of Neural Networks as a Encoder-Decoder with Mutual Information

11 The Stochastic Thermodynamics of Learning

12 Casting Residual Networks as a Mean-Field Optimal Control Problem

References

[Chaudhari et al., 2018] Chaudhari, P., Oberman, A., Osher, S., Soatto, S., and Carlier, G. (2018). Deep relaxation: Partial differential equations for optimizing deep neural networks. *Research in the Mathematical Sciences*, 5(30).

[E et al., 2018] E, W., Han, J., and Li, Q. (2018). A mean-field optimal control formulation of deep learning.

[Fowler, 2018] Fowler, G. (2018). I live with alexa, google assistant and siri. here's which one you should pick. *The Washington Post: The Switch Review*.

[Moskvitch, 2017] Moskvitch, K. (2017). The machines that learned to listen. *BBC Future*.

[Shwartz-Ziv and Tishby, 2017] Shwartz-Ziv, R. and Tishby, N. (2017). Opening the black box of deep neural networks via information.