Overcoming Research Challenges

Three years ago I set a goal for myself: to write a PhD on speech recognition and deep learning. To accomplish this goal, I had to learn a lot, fast.

My approach to learning difficult concepts is the following: teach myself as much as I can, and get help when I hit a wall. As a result, I have become a better learner, better communicator, and more confident researcher. I have learned to formulate my questions well, listen to what is being said, and most importantly, I'm not afraid to fail.

The fear of failure held me back from seeking help for the first year. That year was marked with caffiene fueled late nights in the lab, teaching myself linear algebra, probability theory, and backprop. My progress was slow, so I worked harder. Asking for help didn't seem like an option. However, it became obvious that I couldn't learn everything from books, papers, and blog posts alone. I needed directed answers to my questions. I needed to learn from the experts. The only problem was, I had to go find them.

If the mountain won't come to Mohammed...

My dissertation topic is multi-task learning for deep neural net acoustic modeling in speech recognition for low-resource languages. I needed help in the area I was hitting the most walls: acoustic modeling. Jean-Luc Gauvain and Lori Lamel lead a research team at CNRS near Paris, specilizing on speech recognition for low-resource languages. I reached out to them, asking if I could learn from them as a visiting scholar. They said yes. Together, we applied for an NSF grant and funding from the French Embassy to the US, and were lucky enough to get both. Along with my NSF Graduate Research Fellowship, I had enough funds to work out of their lab for a year. So I packed by bags and headed for Paris.

However, even after I arrived at the lab, I was scared to ask questions. Lori got her PhD from MIT in Electrical Engineering, and Jean-Luc wrote *the* paper on MAP adaptation for GMMs. I felt out of my league. However, there came a day where I had a question that I couldn't figure out, and I knew the people in the next room had the answer. It was a very fundamental question about Expectation-Maximization of Hidden Markov Models. I understood the iterative algorithm (Baum-Welch) after it got started, but I couldn't wrap my head around the very first step: parameter initialization.

I made sure I could express the problem well, drawing out a detailed diagram, and then went over to ask Lori and Jean-Luc. After some back-and-forth clarifications, I had my answer! Each state of the Markov chain is assigned an equal chunk of the observed data sequence: that's all. This is a very simple solution to a hard problem, and the intuition I got was the following: the initial parameters don't actually matter, because the EM algorithm is guaranteed to produce increasingly better models every iteration, approaching the best model given the data. We could just as well randomly initialize the parameters, and given enough iterations of Baum-Welch our model will still be the best possible model. However, by using the data itself as a first guess, our model will start out with a head-start,

shortening the total number of iterations we need to get the best model.

Later that day, at lunch with the post-docs, I said I was releived that I asked the question and got not just an answer, but an intuition. To my surprise they were interested to hear the details of Jean-Luc's answer, because many of them didn't know. These post-docs worked on another section of the speech recognition pipeline (the language model), and didn't know all the intracacies of acoustic modeling. This taught me that it is possible to do good research without being an expert in every facet of the system pipeline. From that day on, I spent more time talking with the researchers and post-docs, learning more at the lunch table than from my textbooks.

Since then, I've taken every chance I get to learn one-on-one from speech technology and machine learning researchers. I spent a week at the University of Edinburgh learning about neural speech synthesis with Simon King, and an afternoon learning about the Kaldi speech toolkit with Dan Povey's lab group at Johns Hopkins University.

My Research Passion & the Google AI Residency

I firmly believe everyone has the right to access speech technology, no matter what language they speak. This goal is a long way off, but I'm working towards it.

People from wealthy countries have access to life-changing technologies, while people in poorer countries don't. This is not an exaggeration when I say life-changing. During my work building a speech synthesizer for blind people in Kyrgyzstan, I heard multiple times how the lack of audiobooks in Kyrgyz schools hinders students from entering university. Even Google Translate still doesn't have a speech synthesizer for Kyrgyz.

This discrepency comes from the lack of labeled training data from these languages. I beleive it is possible to train acoustic models (ie. neural nets) on small datasets, but currently it's not a priority in mainstream machine learning research. Researchers spend a lot of time working on big datasets, because they are accesible for profitable languages (English, German, French, etc.).

Intuitively, I'm working to make better use of small datasets, because collecting massive datasets for every new classifier is not a viable solution to building speech technology for the world's 6,000 languages.

As a Google AI resident, I will have all the experts to help me accomplish my goals. I want to publish influential work, and get more people caring about low-resource languages. This Google residency is the perfect environment for me to learn from and collaborate with people who are dedicated to their work and will push me to higher acheivement. I see the list of languages in Google Translate without speech recognition or speech synthesis boxes waiting to be checked off, and I want to help make that happen.

Thank you for your time and consideration.