Introduction

The following document outlines an internship for Joshua Meyer at the Mozilla Machine Learning Research group. This internship represents an intersection of the interests of both parties as well as an applied continuation of Joshua's academic research. The internship will take place within the Automatic Speech Recognition group (i.e. DeepSpeech) of Mozilla Research. The internship will focus on efforts to bring modern Deep Neural Network speech recognition to low-resource languages.

Mozilla Research

The Mozilla Machine Learning Research group's overarching goal is to democratize speech technologies for the world. To accomplish this, they have made both the tools (i.e. computer code) and materials (i.e. audio recordings) free and open-source for anyone to download and use. The tools have been designed with an easy-to-use programming interface, such that non-expert programmers can begin working quickly. The code and algorithms are maintained under the DeepSpeech team, and the audio data is collected and curated by the Common Voice team. The current internship will be working with the DeepSpeech team on the algorithms and code-base which is used in production all over the world.

Modern speech technologies only exist for a handful of (mostly) European languages, even though there exist well over 6,000 languages in the world. In order to train a speech recognition system for a new language, the developer must first acquire a wealth of resources for that language. Unfortunately, these resources do not exist for most of the world's languages. Specifically, traditional development pipelines require (1) approximately 2,000 hours of transcribed speech, and (2) a pronunciation dictionary containing every word in the language. At Mozilla Research, they are working on a new approach which completely removes the need of the pronunciation dictionary. This new approach is called end-to-end speech recognition, and it only requires a collection of transcribed speech recordings. However, this collection must be very large (approximately 10,000 hours of audio).

Internship Details

Mozilla's DeepSpeech team is the world's leader in open-source end-to-end speech recognition, and Joshua's internship will be focused on developing techniques for bringing this technology to low-resource languages. The internship will take the form of a mentorship, with Joshua working closely with one of the DeepSpeech team members. This team member will be in contact with Joshua via on-site office time. The Mozilla Foundation is an American legal entity, but the DeepSpeech team is located in Europe. As such, the internship will take place on-site in Paris, France, with several visits to Berlin to meet with other members of the Mozilla Machine Learning Research group. The time-line of the internship is August 15th, 2018 - February 15th, 2019.

Joshua's duties will involve the following:

1. Writing and publishing code for training low-resource languages in Deep-Speech (i.e. Python and TensorFlow)

2. Identifying and solving problems in the code-base

3. Working on server-less versions of DeepSpeech (i.e. based on Tensorflow Mobile) for implementation on smartphones and other small-platform devices

4. Identifying ways to improve the user-interface (that is, the DeepSpeech API)

Mozilla Research is expected to provide the following:

1. Mentorship on the implementation of large-scale user-oriented technologies

2. Guidance on best practices for technology team coordination (including git, GitHub, Slack, etc.)

3. Financial support for required internship-related expenses (e.g. travel to meetings and conferences)

4. Best practices on robustly implementing deep learning algorithms (i.e. convolutional, recurrent, and bi-lateral neural networks)

Preparation for the Workforce

This internship at Mozilla will allow Joshua to acquire the skills needed to enter the global workforce - skills which cannot be acquired in a university setting. To enter the machine learning industry, Joshua must understand problems of implementation and best practices in identifying scalable solutions. During his training in academia, Joshua has learned how to create machine learning programs which address certain problems in low-resource data settings, but implementing those programs for millions of users requires a completely different skill set. Taking software into production at large-scale is what modern companies need, and this internship will afford Joshua the opportunity to master those skills.

Broader Impacts and Intellectual Merit

The broader impacts of this collaboration relate to the democratization of language technologies. Joshua will be working with Mozilla to bring free and open-source speech recognition to millions of native speakers of low-resource languages. These technologies can then be easily integrated into applications for the physically disabled, such as Joshua's previous work with blind speakers of the Kyrgyz language.

The intellectual merit of this collaboration relates to Deep Neural Network algorithms, which have found applications in nearly all areas of the sciences. These algorithms currently require massive amounts of data, and advances in low-resource speech recognition will also impact other fields (e.g. bio-medical imaging, object detection) in low-data settings.