Blog of our latest news, updates, and stories for developers

# Kaldi now offers TensorFlow integration

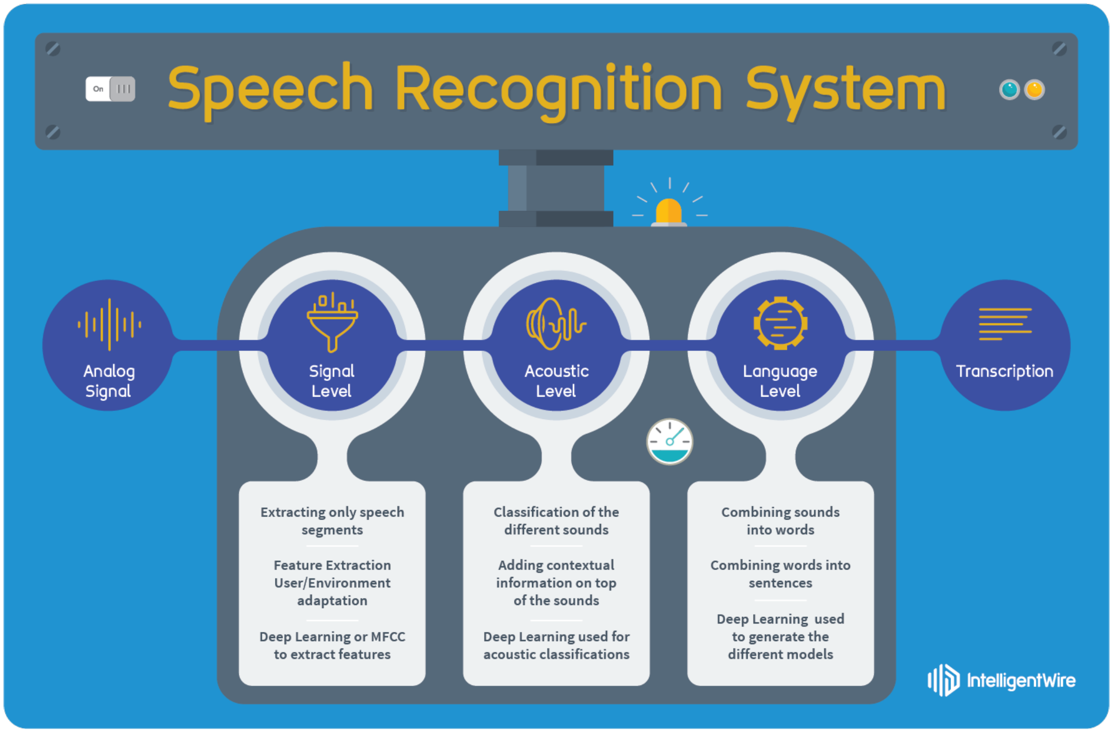
Monday, August 28, 2017

*Posted by Raziel Alvarez, Staff Research Engineer at Google and Yishay Carmiel, Founder of IntelligentWire*

Automatic speech recognition (ASR) has seen widespread adoption due to the recent proliferation of virtual personal assistants and advances in word recognition accuracy from the application of deep learning algorithms. Many speech recognition teams rely on Kaldi, a popular open-source speech recognition toolkit. We're announcing today that Kaldi now offers TensorFlow integration.

With this integration, speech recognition researchers and developers using Kaldi will be able to use TensorFlow to explore and deploy deep learning models in their Kaldi speech recognition pipelines. This will allow the Kaldi community to build even better and more powerful ASR systems as well as providing TensorFlow users with a path to explore ASR while drawing upon the experience of the large community of Kaldi developers.

Building an ASR system that can understand human speech in every language, accent, environment, and type of conversation is an extremely complex undertaking. A traditional ASR system can be seen as a processing pipeline with many separate modules, where each module operates on the output from the previous one. Raw audio data enters the pipeline at one end and a transcription of recognized speech emerges from the other. In the case of Kaldi, these ASR transcriptions are post processed in a variety of ways to support an increasing array of end-user applications.


Speech Recognition System
On
Analog Signal
Signal Level
Acoustic Level
Language Level
Transcription
Extracting only speech segments
Feature Extraction User/Environment adaptation
Deep Learning or MFCC to extract features
Classification of the different sounds
Adding contextual information on top of the sounds
Deep Learning used for acoustic classifications
Combining sounds into words
Combining words into sentences
Deep Learning used to generate the different models
IntelligentWire

Yishay Carmiel and Hainan Xu of Seattle-based IntelligentWire, who led the development of the integration between Kaldi and TensorFlow with support from the two teams, know this complexity first-hand. Their company has developed cloud software to bridge the gap between live phone conversations and business applications. Their goal is to let businesses analyze and act on the contents of the thousands of conversations their representatives have with customers in real-time and automatically handle tasks like data entry or responding to requests. IntelligentWire is currently focused on the contact center market, in which more than 22 million agents throughout the world spend 50 billion hours a year on the phone and about 25 billion hours interfacing with and operating various business applications.

For an ASR system to be useful in this context, it must not only deliver an accurate transcription but do so with very low latency in a way that can be scaled to support many thousands of concurrent conversations efficiently. In situations like this, recent advances in deep learning can help push technical limits, and TensorFlow can be very useful.

In the last few years, deep neural networks have been used to replace many existing ASR modules, resulting in significant gains in word recognition accuracy. These deep learning models typically require processing vast amounts of data at scale, which TensorFlow simplifies. However, several major challenges must still be overcome when developing production-grade ASR systems:

- **Algorithms** - Deep learning algorithms give the best results when tailored to the task at hand, including the acoustic environment (e.g. noise), the specific language spoken, the range of vocabulary, etc. These algorithms are not always easy to adapt once deployed.

- **Data** - Building an ASR system for different languages and different acoustic environments requires large quantities of multiple types of data. Such data may not always be available or may not be suitable for the use case.
- **Scale** - ASR systems that can support massive amounts of usage and many languages typically consume large amounts of computational power.

One of the ASR system modules that exemplifies these challenges is the language model. Language models are a key part of most state-of-the-art ASR systems; they provide linguistic context that helps predict the proper sequence of words and distinguish between words that sound similar. With recent machine learning breakthroughs, speech recognition developers are now using language models based on deep learning, known as neural language models. In particular, recurrent neural language models have shown superior results over classic statistical approaches.

However, the training and deployment of neural language models is complicated and highly time-consuming. For IntelligentWire, the integration of TensorFlow into Kaldi has reduced the ASR development cycle by an order of magnitude. If a language model already exists in TensorFlow, then going from model to proof of concept can take days rather than weeks; for new models, the development time can be reduced from months to weeks. Deploying new TensorFlow models into production Kaldi pipelines is straightforward as well, providing big gains for anyone working directly with Kaldi as well as the promise of more intelligent ASR systems for everyone in the future.

Similarly, this integration provides TensorFlow developers with easy access to a robust ASR platform and the ability to incorporate existing speech processing pipelines, such as Kaldi's powerful acoustic model, into their machine learning applications. Kaldi modules that feed the training of a TensorFlow deep learning model can be swapped cleanly, facilitating exploration, and the same pipeline that is used in production can be reused to evaluate the quality of the model.

We hope this Kaldi-TensorFlow integration will bring these two vibrant open-source communities closer together and support a wide variety of new speech-based products and related research breakthroughs. To get started using Kaldi with TensorFlow, please check out the Kaldi repo and also take a look at an example for Kaldi setup running with TensorFlow.

**Labels:** assistant , automatic speech recognition , intelligentwire , speech , TensorFlow

**No comments :**

**Post a Comment**

Enter your comment...

**Comment as:** 小草 (Google) **Sign out**

**Publish** **Preview** □ Notify me

Google · Privacy · Terms