How Has Deep Learning Revolutionized Human Language Technology?

Tao Ma, Ph.D.

Vice President of the Language Technology Group
ASAPP Inc., Mountain View, California, USA

Abstract:

Communication by voice and text, which we refer to as natural language, is a skill that separates humans from all other species. Only humans possess the complete linguistic package. Since the first computer was invented, it's been our dream to interact with computers using spoken and written words. For decades, machine learning approaches in automatic speech recognition (ASR) and natural language processing (NLP) have been based on shallow models such as Gaussian Mixture Models (GMMs) and hidden Markov models (HMM). These approaches use hand-crafted features and models that attempt to integrate knowledge of speech production, perception and linguistics. We often refer to this knowledge as subject matter expertise. Integration of such knowledge has been a cornerstone of signal processing research for decades.

The emergence of deep learning algorithms has created a paradigm shift in which traditional feature extraction has been replaced with multi-level data-driven feature representations that produces superior results. For example, in speech recognition, the first step in this paradigm shift was to replace mainstream acoustic modeling approaches based on GMMs and HMMs with deep neural networks (DNNs). DNNs were then expanded to the language modeling portion of a speech recognition system by replacing the classic N-gram language model with a neural network language model (NNLM) that re-ranks the top hypotheses returned by the acoustic modeling portion of the system. In recent years, more complex systems have emerged that integrate the acoustic model and language model into a unified large neural network that treats speech recognition as a sequence to sequence learning problem. The end-to-end modeling approach also gained popularity in speech synthesis, with Google's end-to-end speech synthesis system known as Tacotron demonstrating superior prosody and audio fidelity over traditional systems.

In the area of natural language processing, deep learning approaches have achieved widespread performance improvements by embracing data-driven vector embeddings to represent words. Further, new mechanisms such as attention and architectures such as transformers have demonstrated significant improvement across a majority NLP tasks. These advancements enabled a new generation of technology based on pre-trained language model representations (e.g., BERT and GPT-3) that leverage huge amounts of training data using unsupervised learning.

In this talk, we will review how these technologies are transforming traditional contact center applications into intelligent customer interaction platforms. With products like Google Contact Center AI and Amazon Connect, this industry is primed to become the next big thing in language interface innovations. We will introduce some unique approaches to building self-learning engines. We will demonstrate how jointly optimizing ASR and NLP can augment human agents and automate workflows. Consumers will, no doubt, be the ultimate winner in this revolution as customer service will be significantly more efficient, reduce handle time and increase satisfaction.

Biography:

Tao Ma, Ph.D. is the Vice President of Language Technology Group at ASAPP (*https://www.asapp.com/*) where he's leading speech, natural language processing, machine learning, and data curation R&D efforts supporting a portfolio of cutting-edge enterprise AI products. In the past, Dr. Ma has worked at JD.com as

Director of Speech, providing speech recognition and synthesis technologies for automating customer service in call centers, e-commerce shopping apps and smart home devices. Prior to that, Dr. Ma was a Senior Speech Scientist of Siri at Apple. As a founding member of the Siri speech team from 2012/05 to 2017/10, Dr. Ma's work focused on core speech recognition technologies that power Siri, including deep neural networks, statistical language modeling and weighted finite state transducers. Dr. Ma has 15 years of academic and industrial experience across a broad range of speech and natural language technologies. He has previously worked at Intel Labs, Li Creative Technologies, and Telenav. He received a Ph.D. in Computer Engineering from Mississippi State University in 2010.