Invited Talk

Tuesday 27th August 2019

Inductive biases and language emergence in communicative agents

Emmanuel Dupoux

ENS/CNRS/EHESS/INRIA/PSL Research University, Paris

Abstract

Despite spectacular progress in language modeling tasks, neural networks still fall short of the performance of human infants when it comes to learning a language from scarce and noisy data. Such performance presumably stems from human-specific inductive biases in the neural networks sustaining language acquisitions in the child. Here, we use two paradigms to study experimentally such inductive biases in artificial neural networks. The first one relies on iterative learning, where a sequence of agents learn from each other, simulating historical linguistic transmission. We find evidence that sequence to sequence neural models have some of the human inductive biases (like the preference for local dependencies), but lack others (like the preference for non-redundant markers of argument structure). The second paradigm relies on language emergence, where two agents engage in a communicative game. Here we find that sequence to sequence networks lack the preference for efficient communication found in humans, and in fact display an anti-Zipfian law of abbreviation. We conclude that the study of the inductive biases of neural networks is an important topic to improve the data efficiency of current systems.

Short bio

Emmanuel Dupoux directs the Cognitive Machine Learning team at the Ecole Normale Supérieure (ENS) in Paris and INRIA (www.syntheticlearner.com). His education includes a PhD in Cognitive Science (EHESS), a MA in Computer Science (Orsay University) and a BA in Applied Mathematics (Pierre & Marie Curie University, ENS). His research mixes developmental science, cognitive neuroscience, and machine learning, with a focus on the reverse engineering of infant language and cognitive development using unsupervised or weakly supervised learning. He is the recipient of an Advanced ERC grant, the organizer of the Zero Ressource Speech Challenge (2015, 2017, 2019), the Intuitive Physics Benchmark (2019) and led in 2017 a Jelinek Summer Workshop at CMU on multimodal speech learning. He has authored 150 articles in peer reviewed outlets from both cognitive science and language technology.