Inductive Biases for Higher-Level Cognition Deep Learning

Anirudh Goyal

Mila (https://mila.quebec/en/)

More about him to → https://anirudh9119.github.io/

Date & Time: October, 5 (JST) 10:00-11:30

Abstract

One of the challenges right now in reinforcement learning (RL) specifically, AI more generally, is to devise methods for learning complex behaviours that have meaningful generalization outside of the realm of the training distribution. In supervised learning (vision, speech, NLP), it seems that deep models can achieve very complex generalization, but this is not really the case in RL. For an agent interacting with the world, some aspects of the world can be captured using high-level variables, which often have a causal role (referring to objects, and actions or intentions). These high-level variables also seem to satisfy characteristics which low-level data (like pixel level data) do not share. Assumptions on the joint distribution of these high-level variables such as modular decomposition of the knowledge as well as how the distribution changes as a result of intervention may help us to build machine learning systems which can generalize better to out-of-distribution samples. Introducing such assumptions is necessary to even start having a theory about generalizing to out-of-distribution. Many systems consisting of an agent interacting with the world like video games, multiagent modelling etc comprise multiple dynamical processes that operate relatively independently and only occasionally have meaningful interactions. Despite this, most machine learning models employ the opposite inductive bias, i.e., that all processes interact. This can lead to poor generalization (if data is limited) and lack of robustness to changing task distributions. Here, we will talk about inductive biases to factorize knowledge into independent pieces so that they can be combined dynamically and can lead to systematic generalization.

Host

Rei Akaishi, Social Value Decision Making Unit, RIKEN CBS-TOYOTA Collaboration Center (ext.7194)