How to grow a mind: Statistics, structure and abstraction

Joshua B. Tenenbaum
Department of Brain and Cognitive Sciences
MIT

The great puzzle of human learning is this: How do we come to know so much about the world from so little data? From sparse and noisy fragments of experience, we draw generalizations that successfully guide our actions in future situations and tasks we have never faced before. Even young children can infer the meanings of words, the hidden properties of objects, or the existence of causal relations from just one or a few relevant observations -- far outstripping the capabilities of conventional learning machines. How do they do it? And how can we bring machines closer to these human-like learning abilities?

These questions are instances of the classic "problem of induction", with a history going back through Kant and Hume to Plato and Aristotle at the roots of Western philosophy. In a sense, we have had one and only one approach to answering these questions for the last two thousand years. If there are gaps between the data of our experience and what we come to know about the world, some additional knowledge must have already been present, at least implicitly, in order to fill in the gaps. In machine learning, the modern study of induction, this prior knowledge is often called "inductive bias". It is well established that no learning algorithm can succeed without some appropriate form of inductive bias: informative generalizations are not possible at all without constraints on the hypotheses a learner will consider, and the success of generalization depends on how well matched those constraints are to the structure of the learner's environment and tasks.

Recognizing the centrality of prior knowledge, however, is not really an answer to the puzzles of learning. It merely opens the door to the more substantive questions we want to ask. How does background knowledge guide generalization from sparsely observed data? What form does background knowledge take, across different domains and tasks? And how can background knowledge itself be acquired from experience -- how can we learn new ways of understanding the world, or learn to learn? To answer these questions, I will argue, we must bring together several ideas about learning and knowledge that are familiar but have not traditionally played well together. In so doing, we will find these ideas become richer and more powerful than any of them was on its own.

The ideas, in short, are statistics, structure, and abstraction. By "statistics", I mean specifically Bayesian inference in probabilistic generative models. The hypothesis spaces and priors of Bayesian learning provide a natural language in which to describe the inductive biases that guide human generalization. Formalizing these inductive biases, these knowledge-based Bayesian priors, will require us to define probabilistic models over representations that have traditionally been seen as contrasting with (or even opposing) a view of learning as statistical inference: namely, structured symbolic representations such as graphs, grammars, predicate logic, schemas, theories or programs. Finally, to explain the origins of these priors, we will adopt a hierarchical Bayesian framework. Inference occurs in parallel at multiple levels of abstraction, allowing the knowledge that serves as background or inductive constraints for one level of learning to itself be generalized from experience, via simultaneous inferences occurring at higher levels of the hierarchy.

This talk will briefly describe learning and "learning to learn" with hierarchical Bayesian models for categorization, word learning, and causal learning. A key challenge in all of these domains is to balance the need to hold strongly constrained inductive biases -- necessary for rapid generalization -- with the flexibility to adapt to the structure of new environments, learning new inductive biases for which our minds could not have been pre-programmed. Time permitting, I will end with a few comments about how these models of human cognitive growth relate to our current understanding of developmental mechanisms in the brain, and the prospects for using computational models to bridge the cognitive and neural levels of description.