CBMM Memo No. July 18, 2017

Human-like Learning

by

Qianli Liao (and Tomaso Poggio)

Center for Brains, Minds, and Machines, McGovern Institute for Brain Research,
Massachusetts Institute of Technology, Cambridge, MA, 02139.

Abstract:

This work was supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF - 1231216.

1 Introduction

If Machine Learning is a **toolbox**, a *Learning Paradigm* in Machine Learning represents a **tool** (e.g., hammer, wrench, screwdriver, etc.), which largely determines how a machine learning researcher tackle a problem, design experiments and process data. If researchers favor some tools over the others, it greatly influences how they choose their direction towards the goal of stronger AI. Currently, machine learning researchers typically select one or more of the following three learning paradigms to tackle their tasks: Supervised Learning, Unsupervised Learning and Reinforcement Learning.

Supervised Learning directly solves a task by mimicing groundtruth input-target pairs. It heavily relies on **datasets** of such input-target pairs. Example pairs include machine translation sentence pairs, question answering pairs, human dialogue transcripts. Researchers who adopt this paradigm believe that in order to solve a task, we need to first create datasets — by building increasingly larger datasets for all problems, we can solve all of them. Unsupervised Learning typically[1] work similarly to supervised learning but with the target automatically generated by some manually defined criteria, instead of being laboriously annotated.

In contrast, reinforcement learning solves a task by giving learning agent sparse rewards (feedback) with the hope that the agent eventually can learn to perform desired actions. Reinforcement learning typically relies on **environments**. Virtual interactive environments are designed to host and provide reward to machine learning agents. Reinforcement learning can also be performed on "datasets", where rewards are administered based on some actions performed by the agent on the dataset. Researchers in this camp tend to believe that in order to achieve stronger AI, we need to build more realistic environments and perform more simulations with more sophisticated reinforcement learning agents. Again, it is possible that different environments are required for tasks that are different enough (e.g., protein analysis vs. maze navigation).

The above paradigms represent the main "Machine Learning Hammers" that are adopted by most of the researchers. During the last decade, they have brought tremendous successes to the field of AI. Encouraged by the "unreasonable effectiveness" of such "hammers", there seems to be a trend that many researchers start to treat all problems as "nails". The main motivation of this report is to raise the observation that there are many crucial problems in AI (i.e., language, reasoning, knowledge representations, etc.) that are arguably not "nails" — solving them using "hammers" might be acceptable, but there could be better approaches.

In this paper, we propose "Human-like Learning", a paradigm that is sufficiently different from current approaches on the above problems:

We first make the simple observation that all human knowledge has been recorded in natural language. We have a full spectrum of curricula that teaches a person whatever knowledge he/she need to accomplish any task, including learning first language[2], second language, natural science, math, computer science and more. It is a key feature/goal of "Human-like Learning" that we want the AI system to learn from all the raw human knowledge (e.g., over the web, from books). In this sense, this learning paradigm does not really focus on designing "datasets" or "environments" since all recorded human knowledge is **the dataset**. Instead, our paradigm focuses on **systems** — we want to obtain a **minimal** system that can **acquire, store, bootstrap and reason about knowledge like a human**. This system only need to have knowledge representations that are barely enough to bootstrap new knowledge. It does not need even to have seen all English words. But it should be able to learn new words using rather unstructured resources like dictionaries, books and web content. This system can be either **handcrafted ("intelligent design")**, evolved or learned, or obtained by a combination of above approaches. Such a system-centric view of AI is actually reminiscent of "good old school AI" with the exception that we place significant emphasis on : (1) minimal design — use as few rules as possible, but not fewer (2) learnable — it should learn and bootstrap.

There are some advantages from adopting this paradigm:

[1] Of course, there are other forms of unsupervised learning. Here we just describe the most common setting.
[2] Understanding first language is a challenge but we will have special mechanisms for it.

1. **Learning Like Humans With Language:** There are language-based tasks we clearly know how humans learn to solve (i.e., we learn by reading books, tutorials), but current researchers just take exotic approaches (e.g., supervised learning on input-output pairs with some ad-hoc neural models: e.g., seq2seq) because they are predisposed to existing machine learning paradigms. We predict that our proposal will bring more progress to the field in the long-run.

2. **Solve Problems Like Humans With Language:** Assuming learning with natural language can be accomplished, given the plethora of tutorials on any task, this paradigm would allow models to solve any task by reading tutorials, documentations and papers. This would eventually lead to automation of scientific discoveries.

Some people may say that we are just explicitly describing the holy grail of AI research that many people have in mind. But the fact that most people are not working directly on this problem makes our report relevant. To be fair, most researchers are currently adopting existing learning paradigms (e.g., supervised learning and reinforcement learning), playing with complex but add-hoc neural models like LEGO toys, trying to create more datasets, more realistic and complex virtual environments. We simply offer another alternative (or a challenge): can you **design** a minimal system that can bootstrap knowledge like a human from natural materials? Unlike other paradigms, instead of simulating some narrow-domain intelligent behaviours using black-box agents, we focus on: (1) white-box understanding of human intelligence (2) system-level or "meta" learning — representations that learn instead of just learning representations (3) breadth-first strategy: the system should learn to become a generalist before becoming any domain expert.

2 Learning from Examples v.s. Learning from Definition/Description

From another perspective, current machine learning systems mostly rely on learning from examples.

Humans, on the other hand, can learn from merely reading definitions (i.e., descriptions) of things. Examples include reading English dictionaries, reading Wikipedia articles and reading Math textbooks. Learning from definition/description is a key component of the human-like learning we propose.

Of course, ultimately learning should be a combination of all above forms of learning.