CENTER FOR Brains Minds+ Machines

September 22, 2017

Human-like Learning: A Research Proposal [1]

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: We propose Human-like Learning, a new machine learning paradigm aimming at training generalist AI systems in a human-like manner with a focus on human-unique skills.

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

[1]The idea, originated in 2016, has been shared and discussed within lab since 06/2017 (https://github.com/qianli/humanlikelearn).

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 — if we build increasingly larger datasets for all problems, we should eventually be able to solve all of them. Unsupervised Learning typically[2] 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. Similar to creating datasets, 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 some sense, traditional learning paradigms model humans as if they were monkeys who can be trained with only **carrot and stick** in a **trial and error** manner. This treatment is clearly insufficient to support many forms of "human-unique" intelligence like our language skills and scientific understanding.

In this report, 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[3], 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 learned, handcrafted, evolved, 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 hardwiring as possible, but not fewer (2) learnable — it should learn and bootstrap.

[2] Of course, there are other forms of unsupervised learning. Here we just describe the most common setting.

[3] Understanding first language is a challenge but we believe special mechanisms can be designed to tackle it.

There are some advantages of adopting this paradigm:

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., sequence to sequence learning, a.k.a. seq2seq) because they are predisposed to existing machine learning paradigms. For example, when an AI system learns to implement a sorting algorithm, it should just read Wikipedia or an algorithm book to learn it, instead of being trained with input-output pairs of unsorted and sorted numbers (as in supervised learning) or executing random actions and get an "electric shock" whenever it does it wrong (as in reinforcement learning).
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, instead of learning from datasets. This could potentially 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 proposal relevant. To be fair, most researchers are currently adopting existing learning paradigms on highly-specialized tasks, playing with complex but add-hoc neural models, composing neural blocks 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?

We believe that the somewhat unhealthy focus of traditional learning paradigms on "datasets" and "environments" systematically bias the system to narrow-domain AIs. In contrast, the human-like learning we propose is "system-centric" and **"dataset-ignorant"**, which means that we suggest building a "breadth-first" system that gradually learns to handle **all** raw human knowledge (e.g., content from books/web) in a level-wise, simple-to-difficult order — this would more likely lead to a generalist AI instead of a narrow-domain AI — a behavior that is quite different from traditional learning paradigms.

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, since sometimes definition/description could be not available[4].

3 Conclusion

The recent development in AI solved to large extent sensory perception, animal-like, non-language problems using supervised learning and/or reinforcement learning. However, more **human-unique** intelligence problems like language understanding, reasoning, knowledge acquizition, etc. are more challenging for existing learning paradigms. We believe that instead of using traditional learning paradigms (e.g., supervised sequence to sequence learning in many NLP/reasoning tasks as a typical example), a more natural **human-like** learning paradigm

[4] But this scenario is quite rare in most language/science related domains

(e.g., learning by reading books, in a human way) as described in this report is required to make significant progress in these areas.