Radical Empiricism and Machine Learning Research

Judea Pearl

University of California, Los Angeles

Computer Science Department

Los Angeles, CA, 90095-1596, USA

(310) 825-3243 / judea@cs.ucla.edu

Abstract

I contrast the "data fitting" vs. "data interpreting" approaches to data-science along three dimensions: Expediency, Transparency and Explainability. "Data fitting" is driven by the faith that the secret to rational decisions lies in the data itself. In contrast, the data-interpreting school views data, not as a sole source of knowledge but as an auxiliary means for interpreting reality, and "reality" stands for the processes that generate the data." I argue for restoring balance to data science through a task-dependent symbiosis of fitting and interpreting, guided by the Logic of Causation.

Keywords: Causal models, Knowledge representation, Machine learning

Introduction – Simulated Evolution versus Data Science

A speaker at a lecture that I have attended recently summarized the philosophy of machine learning this way: "All knowledge comes from observed data, some from direct sensory experience and some from indirect experience, transmitted to us either culturally or genetically."

The statement was taken as self-evident by the audience, and set the stage for a lecture on how the nature of "knowledge" can be analyzed by examining patterns of conditional probabilities in the data. Naturally, it invoked no notions of "external world," "theory," "data generating process," "cause and effect," "agency," or "mental constructs" because, ostensibly, these notions, too, should emerge from the data if needed. In other words, whatever concepts humans invoke in interpreting data, be their origin cultural, scientific or genetic, can be traced to, and re-derived from the original sensory experience that has endowed those concepts with survival value.

Viewed from artificial intelligence perspective, this data-centric philosophy offers an attractive, if not seductive agenda for machine learning research: In order to develop human level intelligence, we should merely trace the way our ancestors did it, and simulate both genetic and cultural evolutions on a digital machine, taking as input all the data that we can possibly collect. Taken to extremes, such agenda may inspire fairly futuristic and highly ambitious scenarios: start with a simple neural network, resembling a primitive

organism (say an Amoeba), let it interact with the environment, mutate and generate offsprings; given enough time, it will eventually emerge with an Einstein's level of intellect. Indeed, barring sacred scriptures and divine revelation, where else could Einstein acquire his knowledge, talents and intellect if not from the stream of raw data that has impinged upon the human race since antiquities, including of course all the sensory inputs received by more primitive organisms preceding humans.

Before asking how realistic this agenda is, let us preempt the discussion with two observations:

1. Simulated evolution, in some form or another, is indeed the leading paradigm inspiring most machine learning researchers today, especially those engaged in connectionism, deep learning and neural networks technologies which deploy model-free, statistics-based learning strategies. The impressive success of these strategies in applications such as computer vision, voice recognition and self-driving cars has stirred up hopes in the sufficiency and unlimited potentials of these strategies, eroding, at the same time, interest in model-based approaches.[1]

2. The intellectual roots of the data-centric agenda are deeply grounded in the empiricist branch of Western philosophy, according to which sense-experience is the ultimate source of all our concepts and knowledge, with little or no role given to "innate ideas" and "reason" as sources of knowledge (Markie, 2017). Empiricist ideas can be traced to the ancient writings of Aristotle, but have been given prominence by the British empiricists Francis Bacon, John Locke, George Berkeley and David Hume and, more recently, by philosophers such as Charles Sanders Pierce, and William James. Modern connectionism has in fact been viewed as a Triumph of Radical Empiricism over its rationalistic rivals (Buckner, 2019; Lipton, 2015). Indeed, the ability to emulate knowledge acquisition processes on digital machines offer enormously flexible testing grounds in which philosophical theories about the balance between empiricism and innateness can be submitted to experimental evaluation on digital machines.

The merits of testing philosophical theories notwithstanding, I have three major reservations about the wisdom of pursuing a radical empiricist agenda for machine learning research. I will present three arguments why empiricism should be balanced with the principles of model-based science (Pearl, 2019), in which learning is guided by two sources of information: (a) data and (b) man-made models of how data are generated.

I label the three arguments: (1) Expediency, (2) Transparency and (3) Explainability and will discuss them in turns below:

1 Expediency

Evolution is too slow a process (Turing, 1950), since most mutations are useless if not harmful, and waiting for natural selection to distinguish and filter the useful from the useless

---

[1]It will not be an exaggeration to state that societal investment in data fitting technologies, research, education and training have been several thousands times higher than those invested in the model-based science.

is often un-affordable. The bulk of machine learning tasks requires speedy interpretation of, and quick reaction to new and sparse data, too sparse to allow filtering by random mutations. The outbreak of the COVID-19 pandemic is a perfect example of a situation where sparse data, arriving from unreliable and heterogeneous sources required quick interpretation and quick action, based primarily on prior models of epidemic transmission and data production (Pearl, 2020a). In general, machine learning technology is expected to harness a huge amount of scientific knowledge already available, combine it with whatever data can be gathered, and solve crucial societal problems in areas such as health, education, ecology and economics.

Even more importantly, scientific knowledge can speed up evolution by actively guiding the selection or filtering of data and data sources. Choosing what data to consider or what experiments to run requires hypothetical theories of what outcomes are expected from each option, and how likely they are to improve future performance. Such expectations are provided, for example, by causal models that predict both the outcomes of hypothetical manipulations as well the consequences of counterfactual undoing of past events (Pearl, 2019).

2   Transparency

World knowledge, even if evolved spontaneously from raw data, must eventually be compiled and represented in some machine form to be of any use. The purpose of compiled knowledge is to amortize the discovery process over many inference tasks without repeating the former. The compiled representation should then facilitate an efficient production of answers to select set of decision problems, including questions on ways of gathering additional data. Some representations allow for such inferences and others do not. The Ladder of Causation (Pearl and Mackenzie, 2018; Pearl, 2019) defines formally the type of knowledge content needed to answer questions about hypothetical interventions and/or explanations and counterfactuals.

Knowledge compilation involves both abstraction and re-formatting. The former allows for information loss (as in the case of graphical models summarizing numerical equations), while the latter retains the information content and merely transform some of the information from implicit to explicit representations. A classic example would be the spectral representation of a signal wave form; the former is information-equivalent to the latter, but explicitly represent certain aspects of the latter.

These considerations demand that we study the mathematical properties of compiled representations, their inherent limitations, the kind of inferences they support, and how effective they are in producing the answers they are expected to produce. In more concrete terms, machine learning researchers should engage in what is currently called "causal modelling" and use the tools and principles of causal science to guide data exploration and data interpretation processes.

3 Explainability

Regardless of how causal knowledge is accumulated, discovered or stored, the inferences enabled by that knowledge are destined to be delivered to, and benefit a human user. Today, these usages include policy evaluation, personal decisions, generating explanations, assigning credit and blame or making general sense of the world around us. All inferences must therefore be cast in a language that matches the way people organize their world knowledge, namely, the language of cause and effect. It is imperative therefore that machine learning researchers regardless of the methods they deploy for data fitting, be versed in this user-friendly language, its grammar, its universal laws and the way humans interpret or misinterpret the functions that machine learning algorithms discover.[2]

Conclusions

It is a mistake to equate the content of human knowledge with its sense-data origin. The format in which knowledge is stored in the mind (or on a computer) and, in particular, the balance between its implicit vs. explicit components are as important for its characterization as its content or origin.

While radical empiricism may be a valid model of the evolutionary process, it is a bad strategy for machine learning research. It gives a license to the data-centric thinking, currently dominating both statistics and machine learning cultures, according to which the secret to rational decisions lies in the data alone.

A hybrid strategy balancing "data-fitting" with "data-interpretation" better captures the stages of knowledge compilation that the evolutionary processes entails.

References

BUCKNER, C. (2019). Deep learning: A philosophical introduction. *Philosophy Compass* **14** e12625.

LIPTON, Z. (2015). Deep learning and the triumph of empiricism. In *ND Nuggets News*. URL https://www.kdnuggets.com/2015/07/deep-learning-triumph-empiricism-over-theoretical-mathematical-guarantees.html

MARKIE, P. (2017). Rationalism vs. empiricism. In *The Stanford Encyclopedia of Philosophy*, Fall 2017 ed. Metaphysics Research Lab, Stanford University.

PEARL, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Commun. ACM* **62** 54–60. URL http://doi.acm.org/10.1145/3241036

---

[2]Going back to evolutionary perspectives, it is quite possible that human conception of the world is an accidental consequence of catastrophic events, say cosmic radiation, meteorites storms or volcanic eruptions, and will not be discovered by any machine simply because the sense data for such events is not to be found.

PEARL, J. (2020a). Data versus science: Contesting the soul of data-science. In *Causal Analysis in Theory and Practice (Blog)*. July 7.
URL `https://ucla.in/3iEDRVo`

PEARL, J. (2020b). Radical empiricism and machine learning research. In *Causal Analysis in Theory and Practice (Blog)*. July 26.
URL `https://ucla.in/32YKcWy`

PEARL, J. and MACKENZIE, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books, New York.

SCHÖLKOPF, B., LOCATELLO, F., BAUER, S., KE, N. R., KALCHBRENNER, N., GOYAL, A. and BENGIO, Y. (2021). Towards causal representation learning.
URL `https://arxiv.org/abs/2102.11107`

TURING, A. (1950). Computing machinery and intelligence. *Mind* **LIX** 433–460.
URL `https://doi.org/10.1093/mind/LIX.236.433`

Addendum

An email exchange with Yoshua Bengio concerning the arguments above can be found on (Pearl, 2020b).

The discussion focused on the role of causal discovery in human understanding of their environment. Whether causal reasoning should be viewed as a variant of traditional machine learning techniques (Schölkopf et al., 2021), perhaps as a special kind of "inductive bias," or the other way around, that machine learning should be viewed as a supplement to causal inference tasks. I am, of course, of the latter opinion, advocating that even in causal discovery tasks, what we know today about causal inference should be used as guidance to discovery. In particular, we know what features of the world would enable or hinder the discovery of any given structure. I summarized it succinctly saying: "Finding a needle in a haystack is difficult, and it is probably impossible if you haven't seen a needle before."

Most ML researchers today have not seen a needle (i.e., a causal model drawing inferences on interventions and counterfactuals); an educational hindrance that needs to be corrected in order to hasten the discovery of the learning principles we aspire to uncover.