Artificial Intelligence 284 (2020) 103286

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Book review

Judea Pearl, Dana Mackenzie, The Book of Why: The New Science of Cause and Effect, Basic Books, 2018.

Judea Pearl, a Turing Award prize winner, is a true giant of the field of computer science and artificial intelligence. The Turing award is the highest distinction in computer science; i.e., the Nobel Prize of computing. To say that his new book with Dana Mackenzie is timely is, in our view, an understatement. Coming from somebody of his stature and being written for a general audience (unlike his previous books), means that the concerns we have held about both the limitations of solely data driven approaches to artificial intelligence (AI) and the need for a causal approach, will finally reach a very broad audience.

We have long been sceptical of the dominant idea that when 'big data' is coupled with sophisticated machine learning algorithms they can provide 'real' intelligence. Many mainstream computer scientists and data scientists have propagated this idea without meeting resistance, resulting in it being eagerly accepted across government and commercial organisations. This has led to a seemingly unstoppable tidal wave of 'big data solutions' and reports on the 'dangers and opportunities of AI' in the mainstream press. We are sold the idea that we will get access to useful solutions without having to do any kind of considered thinking ourselves, and with little real effort or cost. In many respects this data-driven approach to AI was a reaction to the exact opposite approach – that relied exclusively on experts - that drove the (failed) 'first wave' AI revolution of the 1980s: in that wave it was mistakenly assumed that AI could be achieved from expert-provided 'rules' coupled with massive computing power in the resulting rule-based systems.

In Chapter 1, the core message about the need for causal models is underpinned by what Pearl calls "*The Ladder of Causation*", which is then used to orient the ideas presented throughout the book. Pearl's ladder of causation suggests that there are three steps to achieving true AI. The first step concerns what can be learnt solely from observational data and Pearl argues that we can only learn statistical **associations**. For example, from data we can answer questions like "*what disease best explains the observed symptoms?*" But to be able to answer questions about **interventions** (e.g. "*If I take this drug will it stop the symptoms*?") and **counterfactuals** (e.g. "*If I hadn't taken this drug would my symptoms disappear*?") then we must consider causal models. Answering these essentially causal questions moves us beyond mere association and hence interventions and counterfactuals form the next two steps in Pearl's ladder of causation. Pearl also characterises these three steps on the ladder as 1) 'seeing'; 2) 'doing'; and 3) 'imagining'.

The chapters that follow up to and including Chapter 6 cover interesting historical events, from the birth of statistics to the genesis of causal inference, with many excellent examples and eloquent paradoxes, demonstrating the uncertainty and difficulty in establishing causal relationships from observational data. According to Pearl, the state of the art in AI today is merely a 'souped-up' version of what machines could already do a generation ago: find hidden regularities in a large set of data. "*All the impressive achievements of deep learning amount to just curve fitting*", he said recently.

One of the reasons 'deep learning' has been so successful is that many problems can be solved by optimisation alone without the need to even consider advancing to rungs in the ladder of causation beyond the first. These problems include machine vision and machine listening, natural language processing, robot navigation, as well as other problems that fall within the areas of clustering, pattern recognition and anomaly detection. Big data in these cases is clearly very important and the advances being made using deep learning are undoubtedly impressive, but Pearl convincingly argues that they are not AI.

It is clear that other problems go beyond prediction and demand answers to questions about intervention and counterfactuals and hence, require advancement up the ladder of causation to rungs two and three respectively. Areas like medicine, criminology, marketing, finance and public policy need answers to what are essentially causal questions. A 'smart data' approach combining data with knowledge-based information, representing the underlying causal or temporal aspects of a problem, must be adopted to achieve more intelligent solutions for risk assessment and decision-making. The Bayesian Network framework, that Pearl developed and pioneered some 40 years ago for causal probabilistic reasoning, offers the capability to build suitable smart models that answer these more difficult questions. The book provides a very convincing demonstration of the need for causal models (our own work in this area has always been heavily influenced by Pearl's work).

https://doi.org/10.1016/j.artint.2020.103286

Chapter 7 goes into the details of rung two; interventions. This chapter presents the do-calculus which in some ways is an alternative to randomized control trials, that can often be impractical and expensive, as a valid method for determining causal effects. This chapter is certainly not as easy to appreciate as earlier chapters without a solid understanding of conditional independence and causal models in general. In brief, what the do-calculus enables us to do is to examine the effect of some intervention on some other effect factor within a causal model, and we can retrieve the desired effect of the intervention by making the intervention itself independent of all its causes. Pearl emphasizes that the lack of a calculus for causal modelling and inference is what has held up progress for so long. He notes that, even though such a calculus (notably Bayesian networks) has been around for a while now, it is still shamefully ignored even by many statisticians.

The third rung of the ladder is presented in Chapter 8. Pearl highlights the importance of counterfactual reasoning by arguing that our ability to compare what happened with what could have happened under some alternative scenario is the result of a causal mind that enables us to comprehend responsibility, blame, regret and credit. A system that is able to answer "*what if I had not taken the drug*", is a system that becomes capable of answering questions of counterfactual reasoning based on unobserved evidence.

An important application of counterfactual reasoning is on estimating missing data values, which is a common issue when working with real-world data. This is perhaps a widely underestimated problem since it is often erroneously assumed that missing values in a dataset can be 'predicted', and thus imputed, by examining other observed values in that dataset. Many such statistical imputations methods exist. Pearl argues that these methods are fundamentally flawed due to being 'model-blind'; "*no methods based only of data* (*rung one*) *can answer counterfactual questions* (*rung three*)".

Among the potential applications of counterfactual reasoning, Pearl includes the examples of ***law*** and ***climate change***. In law he notes that counterfactual reasoning corresponds to the very old and well-known notion of 'but-for causation' for which the Model Penal Code provides a test; for example, if Joe blocks a building's fire exit, and Judy dies in a fire after she could not reach the exit, then Joe is legally responsible for her death even though he did not start the fire. This is because 'but-for' his actions Judy would not have died in the fire. Pearl uses this example to introduce probabilities into the argument and he notes (somewhat ironically) that, despite the central role of such probabilistic reasoning in law, the profession has been too conservative and slow to accept mathematical methods. The climate change application is less convincing despite the fact, as Pearl points out, that it is possible to get large amounts of counterfactual data from the simulated outcomes of climate change models. Although, he is critical of simulation models used in natural and social sciences - and recognises the wider problem of placing trust in computer simulations - Pearl suggests that 'by any normal scientific standards the climate models are strong and compelling evidence'. In our view this is not completely consistent with the message of the rest of the book.

The final chapter of the book closes with some general thoughts on AI and Big Data. Scientific research, as well as day-to-day questions, are driven by cause-and-effect. Does smoking cause lung cancer or is there a gene that determines predisposition to cancer? Does gender cause the wage gap? Will this drug cure my disease or treat my symptoms? Would my company's profit be higher had we followed a different marketing strategy? What is the expected impact of this new policy? Did the judge make the correct decision given the evidence? All these questions are causal. Yet, most will try to answer such questions using methods restricted to identifying associations. While the difference between association and causation is nowadays well understood, what has changed over the last few decades is mainly the way the results are stated rather than the way they are generated. While deep learning has been proven to work well for tasks requiring no causal understanding, our understanding of deep learning is almost completely empirical, and we often cannot explain why it works. Some will argue that transparency is not really needed in those cases, and they may be correct. However, transparency does enable effective communication; not just between humans and machines – but also between machines themselves. Regardless, Pearl convincingly argues through his book that causal questions cannot be answered from data alone without a causal model.

If we have one small gripe about the book it is that the language used sometimes equates the results of the causal models reported with mathematical proof. For example, geneticists might consider a genetic confounder to be inconceivable as a causal mechanism for smoking, but something being difficult to conceive is not mathematically equivalent to impossible.

There is much excellent material in this book but, for us, the two key messages are: 1) "True AI" cannot be achieved by data and curve fitting alone, since causal representation of the underlying problems is also required to answer "what-if" questions, and 2) Randomized control trials are not the only 'valid' method for determining causal effects.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This review was supported by the EPSRC project EP/P009964/1, PAMBAYESIAN: Patient Managed decision-support using Bayes Networks, the ERSRC Fellowship project EP/S001646/1 on *Bayesian Artificial Intelligence for Decision Making under Uncertainty*, and by The Alan Turing Institute in the UK under the EPSRC grant EP/N510129/1.

Norman Fenton [*]
Martin Neil [*]
Anthony Constantinou [*]
Risk and Information Management Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, UK
E-mail address: n.fenton@qmul.ac.uk (N. Fenton)

21 December 2018
Available online 21 April 2020

[*] Corresponding authors.