Five takeaways from Mitchell's Artificial Intelligence: A guide for to thinking humans

Preprint · November 2020

DOI: 10.31219/osf.io/9svk3

1 author:

Tung Ho
Phenikaa University
256 PUBLICATIONS **2,717** CITATIONS

SEE PROFILE

Five takeaways from Mitchell's *Artificial Intelligence: A guide for to thinking humans*

Ho Manh Tung

Ritsumeikan Asia Pacific University

Beppu, Oita, Japan

November 16, 2020

In *Artificial Intelligence: A guide for to thinking humans*, Melanie Mitchell, an AI researcher in the Santa Fe Institute, provides an accessible review of the state-of-the-art AI systems around the world and highlights how these advanced systems differ from human intelligence. To this end, I believe the book has achieved its stated purpose. Mitchell, in lay-people language, demystifies *how and where* each current AI technique, i.e., the convolutional neural network, the deep reinforcement learning, the recurrent neural networks, have succeeded and failed in many tasks: images classifications, self-driving cars, speech-recognition, news and videos suggestions, natural language processing, game playing, etc.

The book is divided into segments corresponding to a major AI challenge. The first part is the background of AI, part II is about computer vision, part III is about game playing, part IV is about natural language processing, and finally, part V, the most interesting part, is about the barriers of meaning—the most difficult AI problem.

There are five things that stand out for me. First, most current AI systems require massive human-labeled datasets. From the visual recognition task to natural language processing, all current AI systems use datasets that are labeled by humans. High-quality datasets are the bread and butter of AI, yet, it seems to me there is still a long-standing problem of defining a philosophy of data (Vuong et al., 2018; Napoletani, Panza, & Struppa, 2011; Wickham, 2014). In other words, given a scientific problem, whether it is solved by humans or AI, a researcher still needs to ask how one ought to define the problem's data, its operationalization procedure, its management, etc. Furthermore, he or she can also ask how data should be situated ontologically in relation to facts, knowledge, and information. Those philosophical problems of data are real and serious but it appears there is no unifying framework for them at the moment.

This leads me to the second takeaway. Most intelligence in the current AI systems can be explained by AI researchers' tuning of the "*hyperparameters,*" i.e., all the aspects that require human set-ups so that the learning process can begin: how many iterations, how many parameters, when to explore new paths, when to keep exploited the tried-and-true options, etc. This is the hardest and often the most lucrative job in this field. And if we take the data perspective, it seems to me whenever an AI system fails to perform, we will find there is a problem with the data or the data structure lurking behind somewhere. I believe an unifying philosophy of data can help clarify problems related to the hyperparameters.

Third, all AI systems will fail in very *unhuman* ways well encounter the long-tailed rare events or unfamiliar contexts. Again, taking the data perspective, we are far away from an adequate understanding of how humans perceive and process data and turn them into knowledge. Thus, the popular portrait, in the media, at least, that AI learning is similar to human learning is misleading. Such sensationalism should be combatted in science communication and journalism: both the academia and the media should abide to the principle of intellectual humility, honesty, and openness (Vuong, 2020; Nosek & Errington, 2020).

Fourth, innovation in developing these AI systems require serendipity, i.e., many innovations in a distant field have to come together. For example, the ImageNet algorithm succeeded because its creator, Fei-Fei Li, by serendipity, stumbled upon 1) the database of English words created by psychologist George Miller, which categorizes nouns into hierarchies of abstraction; and 2) the new service offered by Amazon, Mechanical Turk, which hired human workers to perform easy tasks that are currently too hard for machines: object-labeling in a photo, for example. Thus, similar to the market, serendipity's strategic advantage (Napier & Napier, 2013; Vuong & Napier, 2014) should not be underappreciated in advance in the marketplace of ideas.

Finally, as Mitchell (2019) puts it so succinctly, there is a "*barrier of meaning*" in the development of AI: as smart as they are, AIs do not possess any real understanding. However, there is a tendency of both the developers and the media to overhype, overpromise the capacities of AI. The most vivid example is the case of IBM's Watson,

where the mainstream media and the engineers initially claimed that this system could "*read*" millions of books and articles. The connotation of human understanding in the word "read" is misleading. As of now, IBM's Watson's major contracts have been canceled. In the case of newly developed technologies, public perception is not only important for the technologies themselves but also for the trust in science (Vuong, 2018).

I believe, we need more books, podcasts, and articles like Mitchell (2019)'s book in order to rehalibitate the public conversation about AI. To elaborate, the scientific community needs to do a better job in three areas: 1) demystifying new AI systems; 2) focusing on more the mundane technical problems of reliability/accuracy, intrinsic biases, vulnerability to hacking, etc., rather than the threat of the "superintelligent AI," 3) promoting balanced ethical discussion about both the potential benefits and harms of these systems among all stakeholders.

References

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. London: Penguin UK.

Napier NK, Vuong QH (2013). Serendipity as a strategic advantage?. In Wilkinson (ed) *Strategic Management in the 21st Century* (Vol. 1: The Operational Environment), pp. 175-199. Westport, CT: Praeger/ABC-Clio. DOI: 10.13140/2.1.3311.952.

Napoletani, D., Panza, M., & Struppa, D. C. (2011). Agnostic science. Towards a philosophy of data analysis. *Foundations of Science, 16*(1), 1-20.

Nosek, B. A., & Errington, T. M. (2020). The best time to argue about what a replication means? Before you do it. *Nature, 583*(518-520). doi:https://doi.org/10.1038/d41586-020-02142-6

Vuong, Q. H. (2020). Reform retractions to make them more transparent. *Nature, 582*(149). doi:https://doi.org/10.1038/d41586-020-01694-x

Vuong, Q. H., & Napier, N. K. (2014). Making creativity: the value of multiple filters in the innovation process. *International Journal of Transitions and Innovation Systems, 3*(4), 294-327.

Vuong, Q.-H. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour, 2*(1), 5-5. doi:10.1038/s41562-017-0281-4

Vuong, Q.-H., La, V.-P., Vuong, T.-T., Ho, M.-T., Nguyen, H.-K. T., Nguyen, V.-H., . . . Ho, M.-T. (2018). An open database of productivity in Vietnam's social sciences and humanities for public use. *Scientific Data, 5*(1), 180188. doi:10.1038/sdata.2018.188

Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59*(10), 1-23