Advancing Control in the Era of ML and AI

PRAMOD P. KHARGONEKAR
University of California, Irvine
February 3, 2018

How might the current and expected future advances in machine learning and artificial intelligence lead to new opportunities for the systems and control community? I have been motivated by this timely question to write this brief essay. Right at the start, I would like to state that the discussion below is meant to be comprehensive. Rather, my purpose is to present a perspective and some initial thinking on how the systems and control community could engage in this emerging future more fully.

Machine Learning and Artificial Intelligence Context: Over the last decade, great advances have been made in machine learning (ML) and artificial intelligence (AI). Specifically, deep learning architectures, algorithms and techniques have created powerful tools to learn representations of large volumes of data in multiple layers of representation [1]. These tools appear to be very powerful and versatile in learning complex functions and discovering intricate structures in high-dimensional data. They have shown superior performance in image and speech recognition and are being applied in a wide variety of problems: drug discovery, particle physics, astronomy, and biomedicine. More broadly, confluence of insights and techniques from neuroscience, cognitive science, reinforcement learning (RL), and deep learning (DL) has led to very impressive progress in AI with amazing achievements in championship games and demonstration of human level control by an artificial agent [2, 3]. Increasing computational power (thanks to Moore's Law progress) and availability of large amounts of data have been critical to many of these successes and will be increasingly even more important to continuing progress [4, 5]. More importantly, there is a flood of interest, and corresponding investments, from the academic, industrial and government sectors in ML and AI [6].

Control Systems: Control systems have a deep and broad base of foundational knowledge developed over the last 60 years. Dynamic systems modeling, structural properties, identification, stability, feedback, optimality, robustness, fault tolerance, and architecture have been among the central concerns on the theoretical side. These issues have been explored in a wide variety of settings: linear, nonlinear, stochastic, hybrid, distributed, supervisory, and others. Applications have been wide ranging: aerospace, automotive, manufacturing, chemical process, energy, power, transportation, etc. While there is a very rich history, the future is just as promising as there are a multitude of directions for future theoretical and applications research [7].

Future I believe that we are at a truly opportune moment to develop a forward looking vision that can inspire talented researchers for the next decade or more. On the one hand, a major goal of AI is to build machines that can learn and think for themselves [8], including having imagination, reasoning, planning, etc. On the other hand, we have a rich body of knowledge in control systems. The field of control can both benefit from and influence the ongoing revolutionary advances in ML and AI. These advances in ML and AI are going to be driven by the huge increases in computation and data, intense interest, and large investments in these fields across academic, industrial and government entities. By leveraging these ongoing and fortuitous advances in ML/AI, we can aim to have significantly more powerful and versatile control systems. For this, we would need to define specific goals that are currently unachievable with existing control techniques but could potentially be achieved by leveraging ML/AI advances. Such goals would likely be driven by major application areas for control. They would have implications and opportunities for theoretical developments in control. On the other side, we can identify ideas, tools and techniques from control systems that have the potential to advance AI in its quest of building machines that learn and think for themselves.

Of course, there are historic connections between RL [9] and stochastic control [10]. There are more recent connections between sensorimotor neural systems [11, 12], and more generally free energy principle and a unified brain theory [13], and control and estimation which have the potential to advance AI. In a more speculative longer-term direction, there is acceptance within the AI community that rich internal models are critical to human like learning and decision making and that the learning processes must be informed and constrained by prior knowledge. But then there is considerable debate within the AI community on whether such internal models should be configured by human designers or should be learned by the AI agent [8, 14].

Potential Research Directions:

- Traditionally, control systems analysis and design has been based on detailed mathematical models of the system and the environment. These models are typically described using differential equations, discrete-event formalisms, Markov processes, etc. Construction of such models requires highly specific scientific and engineering knowledge, data, and domain expertise. By contrast, ML and (some) AI methods aim to learn models (and control actions) directly from data and experiments. Clearly, in areas where detailed traditional control-oriented models are feasible and have already been developed, there is some scope for ML and AI techniques such as use of deep neural networks for function approximation or rules in discrete-event systems. However, a much larger opportunity arises in areas where such detailed mathematical models do not exist, where performance goals are described at a high level, where the amount of uncertainty is significantly greater, or where the control goals and tasks have high diversity. In such contexts, how can we rethink and re-conceptualize the role of models in control systems in light of what has been learnt in ML and AI in recent years? The big opportunity here for the research community is to open up new formulations where background knowledge from control might be creatively mixed with new paradigms in ML and AI to open new application domains or extend well beyond current performance objectives. Application domains for systems and control are numerous and diverse. Thus, the potential for future research along this line of thinking is very high.
- Neuroscience and cognitive science insights have been key drivers in certain major breakthroughs in AI [3]. The key goal there has been to build machines that can learn and think for themselves [8]. Historically, cybernetics was conceived by Norbert Wiener [15] as "the scientific study of control and communication in the animal and the machine". Over time, this connection between control and cybernetics did not develop as fully as the mathematical control theory paradigm [16]. Can we leverage the new insights at the confluence of neuroscience, cognitive science, reinforcement learning, and AI to conceptualize new architectures for versatile, intelligent and adaptive controllers that work across large diverse domains with improving performance while maintaining safety? Here the big opportunity is to go significantly beyond existing frameworks and paradigms for adaptive control and realize the vision behind Wiener's original cybernetics vision.
- Recent work in AI has led to very impressive results on "human-level control" using an artificial agent that incorporates reinforcement learning and deep Q-network on a large variety of video games [2]. A key challenge here is combine high-dimensional sensory inputs into learning control actions. Thus, the ability of the artificial agent to achieve performance that exceeds all prior algorithms and a level that is comparable to professional human tester is really a control achievement. This advance in AI creates an opportunity to examine the analysis and design of such artificial agents from a control theory perspective. A close collaboration between reinforcement learning, artificial intelligence, and control theory communities might lead to important advances in theory as well as applications.

- While there have been impressive advances in deep learning, many aspects remain only partially understood. For example, some recent results show that deep neural networks easily fit *random* labels [17]. In addition, regularization also does not fully explain the performance of deep neural networks. Thus, there is considerable gap in understanding why deep neural networks have small generalization error in may real world applications. Can methods and tools from systems and control theory offer new analytical perspectives and understanding of this empirical fact? In a related direction, it is now well-know that saddle points are the real barrier in optimization and training of deep neural networks. There is thus an opportunity for systems and control theory community to contribute innovative non-convex optimization solutions to this saddle point problem.
- It is known that many of the machine learning algorithms are not *robust* [18]. For example, image recognition algorithms using deep neural networks can lead to wrong classification if the image is altered in even small ways. This lack of robustness is potentially a major problem, especially if there are adversaries who intend to sabotage intelligent autonomous systems [19]. What are the opportunities to use insights from robust control to increase robustness of these new machine learning algorithms.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

2. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

3. D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.

4. T. Simonite. (2017) How AI Can Keep Accelerating After Moore's Law. [Online]. Available: https://www.technologyreview.com/s/607917/how-ai-can-keep-accelerating-after-moores-law/

5. J. Dean, D. Patterson, and C. Young, "A new golden age in computer architecture: Empowering the machine learning revolution," *IEEE Micro*, 2018.

6. J. Bughin, E. Hazan, S. Ramaswamy, M. Chui, T. Allas, P. Dahlström, N. Henke, and M. Trench. (2017) Artificial intelligence–the next digital frontier. [Online]. Available: https://www.mckinsey.de/files/170620_studie_ai.pdf

7. F. Lamnabhi-Lagarrigue, A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. M. Murray, H. Nijmeijer, T. Samad, D. Tilbury, and P. Van den Hof, "Systems & control for the future of humanity, research agenda: Current and future roles, impact and grand challenges," *Annual Reviews in Control*, vol. 43, pp. 1–64, 2017.

8. B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and Brain Sciences*, vol. 40, pp. 1–72, 2017.

9. R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998.

10. D. P. Bertsekas and J. S. Tsitsiklis, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995.

11. D. M. Wolpert, Z. Ghahramani, and J. R. Flanagan, "Perspectives and problems in motor learning," *Trends in cognitive sciences*, vol. 5, no. 11, pp. 487–494, 2001.

12. R. Grush, "The emulation theory of representation: Motor control, imagery, and perception," *Behavioral and brain sciences*, vol. 27, no. 3, pp. 377–396, 2004.

13. K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, p. 127, 2010.

14. M. Botvinick, D. G. Barrett, P. Battaglia, N. de Freitas, D. Kumaran, J. Z. Leibo, T. Lillicrap, J. Modayil, S. Mohamed, N. C. Rabinowitz *et al.*, "Building machines that learn and think for themselves," *Behavioral and Brain Sciences*, vol. 40, 2017.

15. N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 1961, vol. 25.

16. R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in mathematical system theory*. McGraw-Hill New York, 1969.

17. C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

18. C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.

19. M. Hein and M. Andriushchenko, "Formal guarantees on the robustness of a classifier against adversarial manipulation," in *Advances in Neural Information Processing Systems*, 2017, pp. 2263–2273.