Recap

By Sasha Sax

For the prestigious 2023 Martin Meyerson UC Berkeley Faculty Research Lecture, Prof. Jitendra Malik outlined "The Sensorimotor Road to Artificial Intelligence." Malik begins with the origins of natural intelligence some 500 million years ago: the first animals that could move about had an advantage in finding food, and it is for this purpose that animals evolved efficient perception and motor control. It is only in the last million years that we see the trappings of modern human intelligence, such as language, start to appear. "If the history of intelligent life happened in a day, language only appears in the last three minutes," says Malik.

This is in contrast to the recent progress in artificial intelligence, where AI researchers have made tremendous progress on the linguistic aspects, but the development of a robot with similar sensorimotor capabilities remains elusive. "So how do we explain Moravec's paradox: we have language models like GPT-4 that pass the [LSAT] bar exam but no robots that have the sensorimotor intelligence of a 1 year old," asks Malik. Malik attributes this in part to the abundance of text data compared to the scarcity of sensorimotor data, and also to the difficulty of using that data for sensorimotor control.

Malik outlines a research program to address this. He first situates the approach in the two large intellectual traditions of robotics: pattern recognition and optimal control. According to Malik, his approach revisits the classical problem of adaptive control with the modern tools of deep learning and neural networks, and combines this with insights from biology. As a motivating example, he describes a line of work from his lab known as "Rapid Motor Adaptation" (RMA). The flavor of the approach is that you perform some action, see the consequences of that action, and use that discrepancy between the expected outcome (what biologists call the "efference copy") and the actual outcome in order to see what environmental conditions you are in. The robot then learns appropriate adaptive control laws through trial and error (reinforcement learning). Malik illustrates the successful application of this same basic philosophy on a variety of different motor control problems: quadrupedal walking on varied real-world terrains, bipedal locomotion, in-hand object manipulation using a multi-fingered hand, and more. And these policies are all blind—using only their proprioceptive (joint-angle) signal.

Towards the end of his lecture, Malik sketches a path towards human-level sensorimotor intelligence by drawing on lessons learned from colleagues in cognitive science, neuroscience, and child development. He emphasizes the incremental, multimodal, and physical learning processes in young children and demonstrates how to extend RMA to include these properties. For example, by comparing a visual signal against a time-shifted proprioceptive signal, a quadrupedal robot can learn to leverage RGB cameras with only a few minutes of real-world data. The idea being that if you see an obstacle in front of you, and hit it a second later, then you know there really was an obstacle.

The talk is simultaneously erudite, intuitive and clear, and outlines a compelling counterpoint to what Malik jokingly calls the “tyranny of linguistic imperialism”—a sensorimotor road to artificial intelligence.