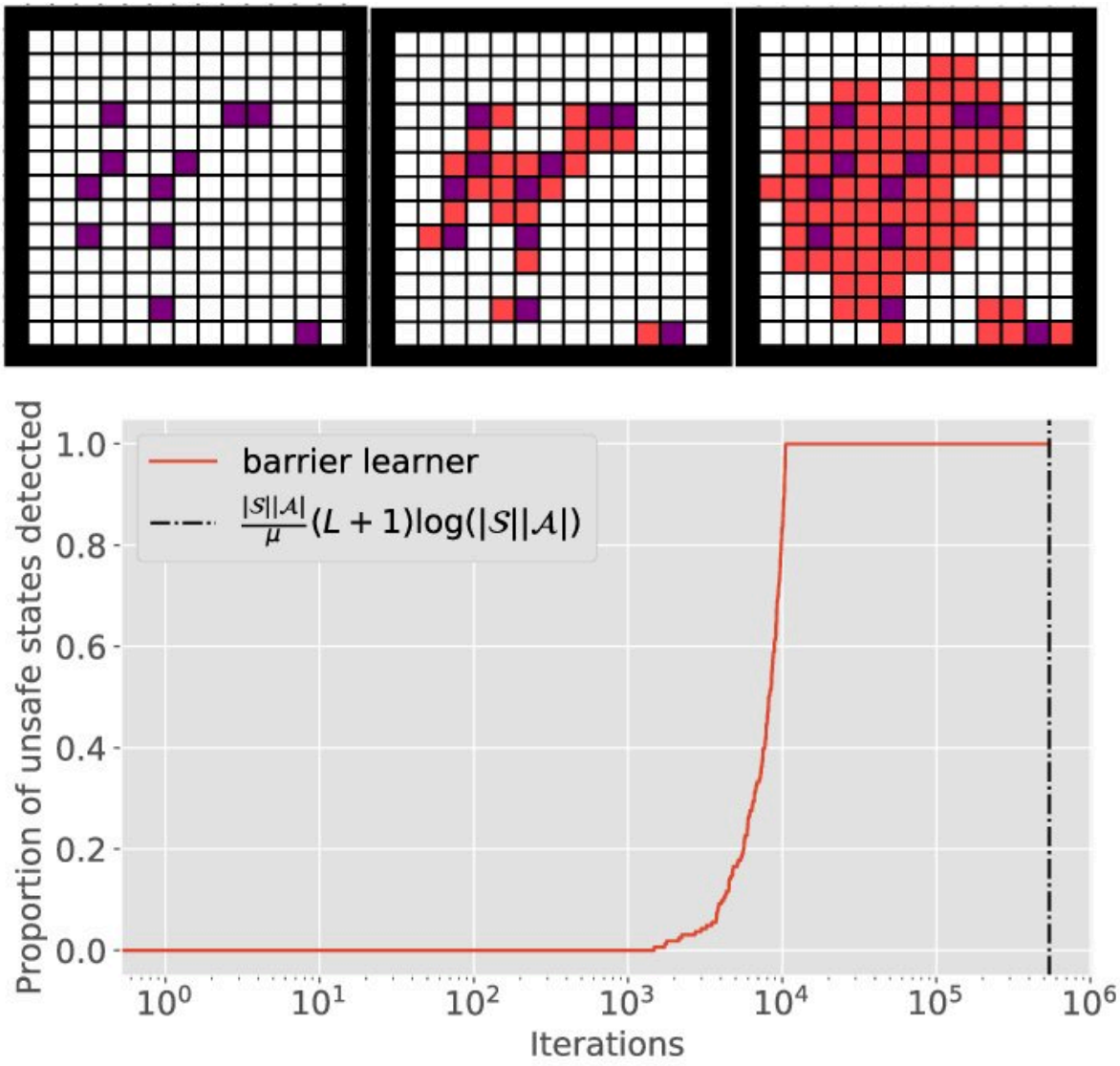
Engineering safer machine learning

June 14 2023, by Maggie Lindenberg



Normalized final exposure for μ = 0.1 as a function of , for varying tolerance level α. Larger values of α and achieve lower handicap (which implies faster detection). Credit: *IEEE Transactions on Automatic Control* (2023). DOI:

10.1109/TAC.2023.3240925

Children first learning to walk may go a bit too fast and fall down, or run into a piece of furniture. However, that cause-and-effect element teaches them invaluable information about how their bodies move through space so that they can avoid falling in the future.

Machines learn in a lot of the same ways that humans do, including learning from their mistakes. However, for many machines—like self-driving cars and power systems—learning on the job with human safety at stake presents a problem. As machine learning matures and proliferates, there is a growing interest in applying it to highly complex, safety-critical autonomous systems. The promise of these technologies, however, is hindered by the safety risks inherent in the training process and beyond.

A new research paper challenges the idea that you need an unlimited number of trials to learn safe actions in unfamiliar environments. The paper, published recently in the journal *IEEE Transactions on Automatic Control*, presents a fresh approach that ensures learning safe actions with complete confidence, while managing the balance between being optimal, encountering dangerous situations, and quickly recognizing unsafe actions.

"Generally, machine learning looks for the most optimized solution, which can result in more errors along the way. That's problematic when the error could mean crashing into a wall," explained Juan Andres Bazerque, assistant professor of electrical and computer engineering at the Swanson School of Engineering, who led the research along with Associate Professor Enrique Mallada at Johns Hopkins University.

"In this study, we show that learning safe policies is fundamentally different from learning optimal policies, and that it can be done separately and efficiently."

The research team conducted studies in two different scenarios to illustrate their concept. By making reasonable assumptions about exploration, they created an algorithm that detects all unsafe actions within a limited number of rounds. The team also tackled the challenge of finding optimal policies for a Markov decision process (MDP) with almost sure constraints.

Their analysis emphasized a tradeoff between the time required to detect unsafe actions in the underlying MDP and the level of exposure to unsafe events. MDP is useful because it provides a mathematical framework for modeling decision-making in situations where outcomes are partly random and partly under the control of a decision maker.

To validate their theoretical findings, the researchers conducted simulations that confirmed the identified tradeoffs. These findings also suggested that incorporating safety constraints can expedite the learning process.

"This research challenges the prevailing belief that learning safe actions requires an unlimited number of trials," stated Bazerque. "Our results demonstrate that by effectively managing tradeoffs between optimality, exposure to unsafe events, and detection time, we can achieve guaranteed safety without an infinite number of explorations. This has significant implications for robotics, autonomous systems, and artificial intelligence, and more."

More information: Agustin Castellano et al, Learning to Act Safely With Limited Exposure and Almost Sure Certainty, *IEEE Transactions on Automatic Control* (2023). DOI: 10.1109/TAC.2023.3240925

Provided by University of Pittsburgh