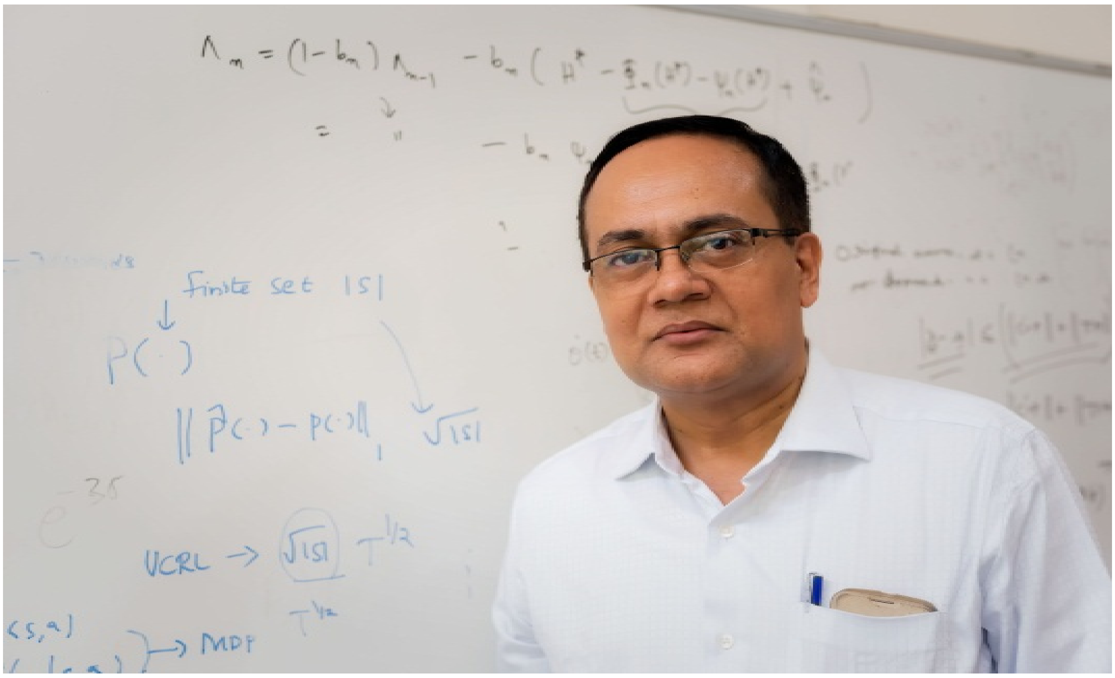
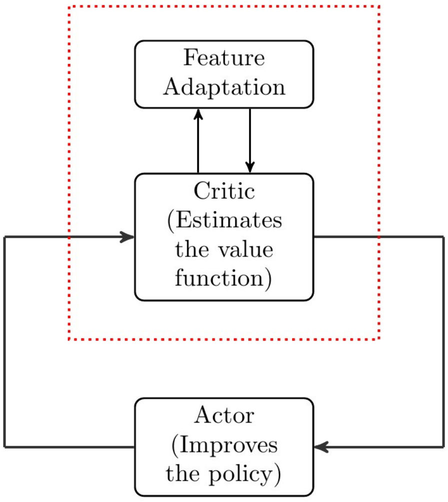
SHALABH BHATNAGAR (Professor, Department of Computer Science and Automation)

MAKING MACHINES LEARN BETTER



Computer programmes that play chess do so not by evaluating every possibility—there are simply far too many—but by looking for indicative "features" in the state of the game (for instance, is the Queen still alive?) This information is then used to assess the "value" of a given state of the system, the objective being to move the system towards a higher value while minimising "cost" (is it worth losing a pawn to protect the King?). Real-life optimization problems are much like a game of chess—complex.

ShalabhBhatnagar and his group work on the theory and applications of reinforcement learning (RL), an area of machine learning used in optimization problems. They are particularly interested in traffic-handling—the vehicular kind on our roads, and the digital kind in our wireless networks. In RL, the algorithm "learns" with experience by modifying, over time, the features it uses to capture the state of the system. This allows for the optimization even of stochastic systems. These "actor-critic" algorithms can be used to maintain average costs over a given time period, or to minimize long-term costs.



These "actor-critic" algorithms can be used to maintain average costs over a given time period, or to minimize long-term costs

Back in 2013, Bhatnagar and his former student, KJ Prabuchandran, now working for IBM Research, in collaboration with VivekBorkar, an IIT Bombay professor, devised an algorithm that uses this kind of "feature adaptation" to optimize systems of interest. They showed that the error resulting from the feature adaptation disappears in the limit and one recovers the true value function, i.e., the one obtained from full system knowledge. In a more recent study, they improved their algorithm by allowing it to change the action it would take for a given state of the system.*They parameterize the action that the "actor" part of their algorithm takes; this parameter can then be varied as the algorithm learns from experience. The "critic" part of the algorithm evaluates the action taken using a features-based value estimation as before.

* KJ Prabuchandran, ShalabhBhatnagar, and Vivek S Borkar. 2016. Actor-critic algorithms with online feature adaptation. *ACM Trans. Model.Comput.Simul*. 26 (4): Article 24.