CRUNCH Seminars at Brown, Division of Applied Mathematics

Friday - August 7, 2020

Discovering Reinforcement Learning Algorithms

Sumit Vashishtha

Reinforcement learning (RL) algorithms update an agent's parameters according to one of several possible rules, discovered manually through years of research. Automating the discovery of update rules from data could lead to more efficient algorithms, or algorithms that are better adapted to specific environments. Although there have been prior attempts at addressing this significant scientific challenge, it remains an open question whether it is feasible to discover alternatives to fundamental concepts of RL such as value functions and temporal-difference learning. This paper introduces a new meta-learning approach that discovers an entire update rule which includes both 'what to predict' (e.g. value functions) and 'how to learn from it' (e.g. bootstrapping) by interacting with a set of environments. The output of this method is an RL algorithm that we call Learned Policy Gradient (LPG). Empirical results show that our method discovers its own alternative to the concept of value functions. Furthermore it discovers a bootstrapping mechanism to maintain and use its predictions. Surprisingly, when trained solely on toy environments, LPG generalises effectively to complex Atari games and achieves non-trivial performance. This shows the potential to discover general RL algorithms from data.