Seeking entropy: complex behavior from intrinsic motivation to occupy action-state path space

Jorge Ramírez-Ruiz[1], Dmytro Grytskyy[1], and Rubén Moreno-Bote[1,2]

[1]Center for Brain and Cognition, and Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
[2]Serra Húnter Fellow Programme, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Intrinsic motivation generates behaviors that do not necessarily lead to immediate reward, but help exploration and learning. While there are many efforts to study intrinsic motivation, it is usually studied in a reward maximization framework. Here we show that agents having the sole goal of maximizing occupancy of future actions and states, that is, moving and exploring on the long term, are capable of complex behavior without any reference to extrinsic rewards. We find that action-state path entropy is the only measure consistent with additivity and other intuitive properties of expected future action-state path occupancy. We provide analytical expressions that relate the optimal policy with the optimal state-value function, from where we prove uniqueness of the solution of the associated Bellman equation and convergence of our algorithm to the optimal state-value function. Using discrete and continuous state tasks, we show that 'dancing', hide-and-seek and a basic form of altruistic behavior naturally result from entropy seeking without extrinsic rewards. Intrinsically motivated agents can objectively determine what states constitute rewards, exploiting them to ultimately maximize action-state path entropy. These results aim to point towards an understanding of behavior and its variability from the perspective of the internal needs of the agents, as well as to provide an algorithmic procedure to generate local stochastic behavior that efficiently explores the environment while respecting the ecological constraints of the agent.

Acknowledgments

This work is supported by the Howard Hughes Medical Institute (HHMI, ref 55008742) and MINECO (Spain; BFU2017-85936-P) to R.M.-B, and MINECO/ESF (Spain; PRE2018- 084757) to J.R.-R. We thank Jose Apesteguia, Luca Bonatti, Ignasi Cos, Benjamin Hayden and Chiara Mastrogiuseppe for very useful comments.