CNeuro2022 Lecture Abstracts

Ruben Moreno Bote

Abstract 1 – Basic Lecture:

Soft Reinforcement Learning

In RL (reinforcement learning), agents are assumed to maximize some external or internal objective, typically defined as a discounted sum of future rewards, and it has become a very popular tool to solve very hard AI control problems and is a guiding theory in Neuroscience. I will make an introduction to the basic notions of RL, such as reward, value function, Q-value and Bellman equation. Then, I will introduce ways researchers have thought of to favor exploration and smoothness in the action selection, which helps learning in some cases. These methods are mostly based on adding action-entropy terms to regularize the reward objective [Todorov(2009), Peters et al.(2010)]. Then, I will describe how researchers have started to think about eliminating external rewards as objective and defined it in terms of intrinsic motivation signals, such as empowerment [Klyubin et al.(2005)] and causal entropic forces [Wissner-Gross and Freer(2013)]. After commenting on the limitations of those approaches, I will briefly introduce entropy-seeking RL (or Pure MaxEnt RL), based on the hypothesis that natural agents maximize future action-state path occupancy [Ramírez-Ruiz et al.(2022)]. Next lecture will take entropy-seeking RL as starting point.

Abstract 2 – Advanced Lecture:

Entropy-Seeking RL

Intrinsic motivation generates behaviors that do not necessarily lead to immediate reward, but help exploration and learning. I will 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 [Ramírez-Ruiz et al.(2022)]. I will show that action-state path entropy is the only measure consistent with additivity and other intuitive properties of expected future action-state path

occupancy. Using discrete and continuous state tasks, I will finally show that 'dancing', hide-and-seek and a basic form of altruistic behavior naturally result from entropy seeking without external rewards.

References

[Todorov(2009)] Emanuel Todorov. Efficient computation of optimal actions. *Proceedings of the national academy of sciences*, 106(28):11478–11483, 2009.

[Peters et al.(2010)] Jan Peters, Katharina Mulling, and Yasemin Altun. Relative entropy policy search. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[Klyubin et al.(2005)] Alexander S Klyubin, Daniel Polani, and Chrystopher L Nehaniv. Empowerment: A universal agent-centric measure of control. In *2005 ieee congress on evolutionary computation*, volume 1, pages 128–135. IEEE, 2005.

[Wissner-Gross and Freer(2013)] Alexander D Wissner-Gross and Cameron E Freer. Causal entropic forces. *Physical review letters*, 110(16):168702, 2013.

[Ramírez-Ruiz et al.(2022)] Jorge Ramírez-Ruiz, Dmytro Grytskyy, and Ruben Moreno-Bote. Seeking entropy: complex behavior from intrinsic motivation to occupy action-state path space. *arXiv preprint arXiv:2205.10316*, 2022.