Extending the Maximal Occupancy Principle in partially observable scenarios

The field of reinforcement learning has been prominent in recent years. Typically, reinforcement learning aims to maximise an external reward, which has led to impressive achievements [5]. However, many behaviours appear to stem from internal motivations of the agent. Curiosity, exploration, or spontaneous movements are not always well explained by reward maximisation alone.

The Maximal Occupancy Principle (MOP) [4] is an approach that moves away from reward maximisation. Instead, it proposes that behaviour aims to maximise occupancy over future paths of actions and states. While simple in formulation, this approach can generate complex behaviours, combining both variability and goal-directedness without relying on explicit rewards [3]. MOP, as originally proposed, assumes a fully observable environment, where the agent has access to all system variables. In most natural settings, however, agents can only observe part of the environment, often through noisy inputs (e.g., weak smells or distant, unclear objects). In such cases, the agent must form beliefs about hidden variables and estimate the underlying state of the system.

In this work, we extend MOP to partially observable environments. We define the equations for belief propagation over hidden variables and reformulate the value function that governs policy selection. We demonstrate the resulting behaviour in grid world scenarios. We also compare against other agents, including a standard reward-seeking agent (R-agent) and two reward-free approaches: active inference [1] and empowerment [2].

References

[1] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, Giovanni Pezzulo, et al. Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68:862–879, 2016.

[2] 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.

[3] Rubén Moreno-Bote and Jorge Ramirez-Ruiz. Empowerment, free energy principle and maximum occupancy principle compared. In *NeurIPS 2023 workshop: Information-Theoretic Principles in Cognitive Systems*, 2023.

[4] Jorge Ramirez-Ruiz, Dmytro Grytskyy, Chiara Mastrogiuseppe, Yamen Habib, and Ruben Moreno-Bote. Complex behavior from intrinsic motivation to occupy future action-state path space. *Nature Communications*, 15(1):6368, 2024.

[5] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.