The Agent-Environment Boundary

Anonymous Author(s)
Affiliation
Address
email

The standard reinforcement learning (RL) interaction [Sutton and Barto, 2020] (Section 3.1) loop is often depicted as an agent that observes the current state of the environment, interacts with it through actions, and receives a reward at discrete time steps. The typical problem formulation in Markov decision processes (MDP) involves a predefined set of states and actions through which we conceptualize the agent's interaction with the environment. A philosophical question arises when considering the actual boundary between the agent and the environment, particularly in the context of humans or animals. Is the agent the entire human, with everything outside their physical body considered the environment? Or should we adopt a more microscopic view, where the environment includes the human body, and the agent is the brain, which sends neural signals to control the body and interact with the world? Alternatively, a more macroscopic view might consider the agent to be something outside the human body, perhaps associated with a group or collective.

The need for a clear distinction between the agent and the environment arises from the necessity to partition parts of the system that are directly controllable or modifiable by the agent—in this case, the action set [Sutton and Barto, 2020]. This distinction is crucial for influencing variables that may immediately or causally affect the state variables. For instance, in standard robotics tasks, the action space often consists of joint torques (or even the current to the motor joints), as these are directly controllable by the agent to influence the position and motion of limbs via joint orientations. Additionally, the observation space may include external inputs, such as an image feed of the surrounding environment, enabling the agent to achieve its goals or complete tasks.

We aim to propose a more fluid definition of the agent-environment boundary that naturally integrates the idea of actions as part of the observation space. These actions, controllable by the agent at the smallest temporal scale, can influence variables in the observation space over longer time scales or until the agent develops a better understanding of its morphological arrangement to possibly augment its action space with new concepts. This understanding allows the agent to control and predict its observation space, a concept related to building predictive knowledge, such as general value functions (GVFs) [Sutton et al., 2011]. It is at this boundary that the agent needs to have intrinsic tasks defined, enabling the use of RL principles first to learn a causal model of effects. The agent can then build models with varying levels of abstraction, allowing it to plan and perform RL at an extended frontier. This relates strongly to empowerment [Jung et al., 2012, Salge et al., 2013] and constructivism [Drescher, 1991].

Consider the example of a human baby: at first, a baby engages in seemingly random limb movements, much like the random exploration seen in RL, as a way to make sense of their sensory experiences. Over time, as the baby grows into a child and eventually an adult, these movements become more coordinated and purposeful, building on the primitive muscle actions learned in early development. This same principle applies when learning a new, complex skill, like driving a car. On your first day behind the wheel, the mechanics of the vehicle are unfamiliar, and you must consciously control your hands and feet to navigate. At this stage, your body is the agent, and the car is part of the environment you interact with. Fast forward a year: after enough practice, the car feels like an extension of your body, and the agent now encompasses both you and the vehicle. The environment has shifted outward, as you've built a mental model that allows you to predict and control the car with ease.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

References

G. L. Drescher. *Made-up Minds: A Constructivist Approach to Artificial Intelligence*. Artificial Intelligence. MIT Press, Cambridge, Mass., 1991. ISBN 978-0-262-04120-1.

T. Jung, D. Polani, and P. Stone. Empowerment for Continuous Agent-Environment Systems. https://arxiv.org/abs/1201.6583v1, Jan. 2012.

C. Salge, C. Glackin, and D. Polani. Empowerment – an Introduction. https://arxiv.org/abs/1310.1863v2, Oct. 2013.

R. S. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts London, England, second edition edition, 2020. ISBN 978-0-262-03924-6.

R. S. Sutton, J. Modayil, M. Delp, T. Degris, P. M. Pilarski, A. White, and D. Precup. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '11, pages 761–768, Richland, SC, May 2011. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 978-0-9826571-6-4.