Optimal Universal Explorative Agents

Laurent Orseau[1] Tor Lattimore[2]
Marcus Hutter[2]

[1]AgroParisTech, UMR 518 MIA, F-75005 Paris, France
INRA, UMR 518 MIA, F-75005 Paris, France
[2] RSCS, Australian National University
Canberra, ACT, 0200, Australia

Extended Abstract

We present some recent research [Ors11, OLH13] in the area of universal reinforcement learning about optimally explorative agents, which we call knowledge-seeking agents. Their purpose is to gather as much information as possible about their environment, as fast as possible. We are interested in such an optimal strategy and its properties and implications, as the limits of what can be done. To this aim, the model is based on Hutter's AIXI [Hut05], which defines the optimal reinforcement learning agent, for all computable stochastic environments, provided it has access to infinite computation resources.

Studying universal reinforcement learning is informative for several reasons: (a) it allows for a *universal definition* of what we would like to do if we could, (b) getting rid of computational aspects allows us to focus on the core problems of RL, (c) it gives a *target definition* of what we should try to approximate given resource constraints, (d) it allows to consider environments of arbitrary complexity, without being restrained to MDPs or POMDPs (like counting environments), and (e) it gives an upper bound on what is feasible. Regarding the latter point, these models allowed us to prove that even the universal Bayesian RL agent can stop exploring in some situations, leading to suboptimal reward gathering in some environments [Ors13]: Even without computational constraints, the exploration/exploitation dilemma is still present. In turn, this allowed to determine the limitations of what *any* RL agent can do [LH11], and how to make sure it explores its environment while gathering as much reward as it can. In particular, it can be shown that a trade-off really is necessary, and that the agent cannot hope to (asymptotically) collect all the rewards.

Exploitation without exploration makes no sense unless the agent already knows the environment. Exploration without exploitation, on the other hand, is a fundamentally interesting learning problem. This makes even more sense if one considers an agent that is rewarded (internally) for gaining information about its environment: Then exploration *is* exploitation. In 2011, we gave the first universal model for such an agent, focusing on deterministic environments [Ors11], and proved that the agent asymptotically learns everything it *can* learn about its environment. Its definition is surprisingly simple, but alas it fails in stochastic environments. Recently, we proposed a new knowledge-seeking agent that is noise-resistant while still being able to learn its environment, in the sense that it learns to predict the future for all policies. We end up with an optimally explorative agent, with a simple definition, that can be seen as the definition of the optimal scientist, whose goal is to gain as much knowledge as it can about its environment. However, being able to predict all futures is not sufficient in itself to define a scientific behaviour, and one must also make sure that the agent will not intentionally take actions that will make all futures easy to predict, *e.g.* by destroying the world. We proved that our agent will avoid such situations, and in a sense will avoid destroying information if possible.

However, we are still left with the choice of the horizon function (a generalisation of the discount factor), and two other parameters. Requiring a horizon function is unsatisfactory, not only because it is a free parameter but also because it makes the agent somewhat myopic, which means that events in the distant future have less weight, even if they are very probable. Studying universal reinforcement learning can give us insights as to how these free parameters can be removed, or why this is not possible.

References

[Hut05] Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability.* Springer, Berlin, 2005.

[LH11] Tor Lattimore and Marcus Hutter. Asymptotically optimal agents. *Algorithmic Learning Theory*, 6925:368–382, 2011.

[OLH13] L. Orseau, T. Lattimore, and M. Hutter. Universal knowledge-seeking agents for stochastic environments. In *Proc. 24nd International Conf. on Algorithmic Learning Theory (ALT'13).*, LNAI, Singapore, Rep. of Singapore, (To appear) 2013. Springer, Berlin.

[Ors11] Laurent Orseau. Universal Knowledge-Seeking Agents. In Jyrki Kivinen, Csaba Szepesvári, Esko Ukkonen, and Thomas Zeugmann, editors, *Algorithmic Learning Theory (ALT)*, volume 6925 of *LNAI*, pages 353–367, Espoo, Finland, 2011. Springer.

[Ors13] Laurent Orseau. Asymptotic non-learnability of universal agents with computable horizon functions. *Theoretical Computer Science*, 473:149 – 156, 2013.