On the Computability of Solomonoff Induction and Knowledge-Seeking

Jan Leike and Marcus Hutter

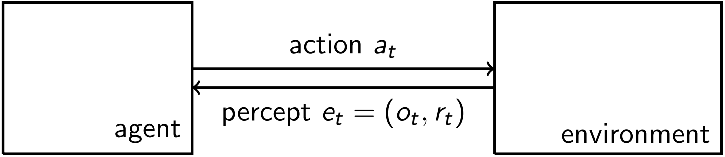
http://jan.leike.name/

Australian National University

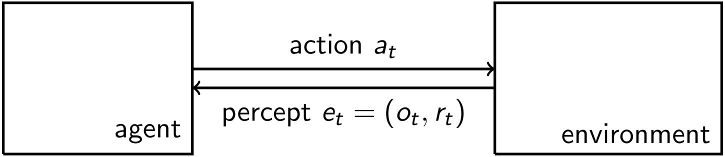
ALT'15 — 6 October 2015

Outline

Reinforcement Learning

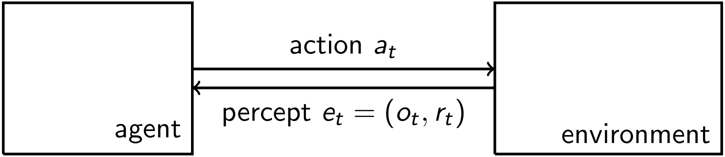


Reinforcement Learning



Goal: maximize $\sum_{t=1}^{\infty} r_t$

Reinforcement Learning

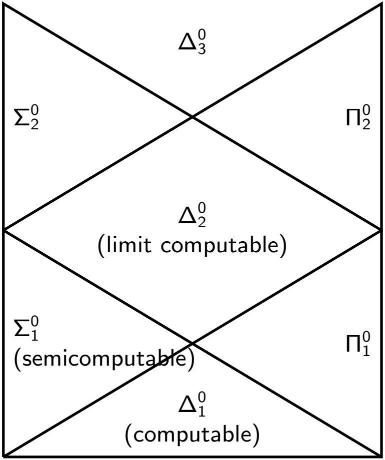


Goal: maximize $\sum_{t=1}^{\infty} r_t$

[Hutter, 2005]:

$$\text{AIXI} = \text{Solomonoff Induction} + \text{Expectimax search}$$

The Arithmetical Hierarchy



Want: limit computable optimal reinforcement learning agent

Computability of Optimal Policies

[Leike and Hutter, 2015b]

Model	Optimal	ε-Optimal
Computable Env	Δ_2^0	Δ_1^0
Semicomputable Env	Δ_3^0, Π_2^0-hard	Δ_2^0, Σ_1^0-hard
AIXI	Δ_3^0, Σ_1^0-hard	Δ_2^0, Σ_1^0-hard

Optimal Agents

- AIXI is not asymptotically optimal [Orseau, 2013]

Optimal Agents

- AIXI is not asymptotically optimal [Orseau, 2013]
- Bad priors cause AIXI to misbehave drastically [Leike and Hutter, 2015a]

Optimal Agents

- AIXI is not asymptotically optimal [Orseau, 2013]
- Bad priors cause AIXI to misbehave drastically [Leike and Hutter, 2015a]
- Weak Asymptotic Optimality [Lattimore and Hutter, 2011]:

$$\frac{1}{t}\sum_{k=1}^{t}(V_k^{\pi\mu} - V_k^{*\mu}) \to 0 \ \mu\text{-}\pi\text{-almost surely as } t \to \infty$$

Optimal Agents

- AIXI is not asymptotically optimal [Orseau, 2013]
- Bad priors cause AIXI to misbehave drastically [Leike and Hutter, 2015a]
- Weak Asymptotic Optimality [Lattimore and Hutter, 2011]:

$$\frac{1}{t}\sum_{k=1}^{t}(V_k^{\pi\mu} - V_k^{*\mu}) \to 0 \ \mu\text{-}\pi\text{-almost surely as } t \to \infty$$

- BayesExp is weakly asymptotically optimal [Lattimore, 2013]

Outline

Computability of Solomonoff Induction

$$M(x) := \sum_{p \colon x \sqsubseteq U(p)} 2^{-|p|}$$

Computability of Solomonoff Induction

$$M(x) := \sum_{p:\, x \sqsubseteq U(p)} 2^{-|p|} \quad M_{\mathrm{norm}}(xa) := M_{\mathrm{norm}}(x) \frac{M(xa)}{\sum_{b \in \mathcal{X}} M(xb)}$$

Computability of Solomonoff Induction

$$M(x) := \sum_{p:\, x \sqsubseteq U(p)} 2^{-|p|} \quad M_{\mathrm{norm}}(xa) := M_{\mathrm{norm}}(x) \frac{M(xa)}{\sum_{b \in \mathcal{X}} M(xb)}$$

$$\overline{M}(x) := \lim_{n \to \infty} \sum_{y \in \mathcal{X}^n} M(xy)$$

Computability of Solomonoff Induction

$$M(x) := \sum_{p:\, x \sqsubseteq U(p)} 2^{-|p|} \qquad M_{\mathrm{norm}}(xa) := M_{\mathrm{norm}}(x) \frac{M(xa)}{\sum_{b \in \mathcal{X}} M(xb)}$$

$$\overline{M}(x) := \lim_{n \to \infty} \sum_{y \in \mathcal{X}^n} M(xy) \qquad \overline{M}_{\mathrm{norm}}(x) := \frac{\overline{M}(x)}{\overline{M}(\epsilon)}$$

Computability of Solomonoff Induction

$$M(x) := \sum_{p:\, x \sqsubseteq U(p)} 2^{-|p|} \quad M_{\text{norm}}(xa) := M_{\text{norm}}(x) \frac{M(xa)}{\sum_{b \in \mathcal{X}} M(xb)}$$

$$\overline{M}(x) := \lim_{n \to \infty} \sum_{y \in \mathcal{X}^n} M(xy) \qquad \overline{M}_{\text{norm}}(x) := \frac{\overline{M}(x)}{\overline{M}(\epsilon)}$$

P	$P(x) > q$	$P(xy \mid x) > q$
M	$\Sigma^0_1 \setminus \Delta^0_1$	$\Delta^0_2 \setminus (\Sigma^0_1 \cup \Pi^0_1)$
M_{norm}	$\Delta^0_2 \setminus (\Sigma^0_1 \cup \Pi^0_1)$	$\Delta^0_2 \setminus (\Sigma^0_1 \cup \Pi^0_1)$
$\overline{M}$	$\Pi^0_2 \setminus \Delta^0_2$	$\Delta^0_3 \setminus (\Sigma^0_2 \cup \Pi^0_2)$
$\overline{M}_{\text{norm}}$	$\Delta^0_3 \setminus (\Sigma^0_2 \cup \Pi^0_2)$	$\Delta^0_3 \setminus (\Sigma^0_2 \cup \Pi^0_2)$

Outline

Information-Seeking and Entropy-Seeking

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ is the history
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Information-Seeking and Entropy-Seeking

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ is the history
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Entropy-Seeking Value Function [Orseau, 2014]:

$$V_H^\pi(æ_{<t}) := \sum_{e_{t:m}} -\xi_{\mathrm{norm}}(e_{1:m} \mid e_{<t} \parallel a_{1:m}) \log \xi_{\mathrm{norm}}(e_{1:m} \mid e_{<t} \parallel a_{1:m})$$

where $a_i := \pi(e_{<i})$ for all $i \geq t$.

Information-Seeking and Entropy-Seeking

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ is the history
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Entropy-Seeking Value Function [Orseau, 2014]:

$$V_H^\pi(æ_{<t}) := \sum_{e_{t:m}} -\xi_{\text{norm}}(e_{1:m} \mid e_{<t} \parallel a_{1:m}) \log \xi_{\text{norm}}(e_{1:m} \mid e_{<t} \parallel a_{1:m})$$

where $a_i := \pi(e_{<i})$ for all $i \geq t$.

Information-Seeking Value Function [Orseau et al., 2013]:

$$V_I^\pi(æ_{<t}) := \sum_{e_{t:m}} \sum_{\nu \in \mathcal{M}} w_\nu \frac{\nu(e_{1:m} \| a_{1:m})}{\xi_{\text{norm}}(e_{<t} \| a_{<t})} \log \frac{\nu(e_{1:m} | e_{<t} \| a_{1:m})}{\xi_{\text{norm}}(e_{1:m} | e_{<t} \| a_{1:m})}$$

Knowledge-Seeking is Limit Computable

Model	Optimal	ε-Optimal
Entropy-Seeking	Δ_3^0	Δ_2^0
Information-Seeking	Δ_3^0	Δ_2^0

(for finite horizon)

Outline

Limit-Computable BayesExp

Effective horizon:

$$H_t(\varepsilon) := \min \left\{ k \mid \frac{\sum_{i=t+k}^{\infty} \gamma(i)}{\sum_{i=t}^{\infty} \gamma(i)} \leq \varepsilon \right\}$$

Limit-Computable BayesExp

Effective horizon:

$$H_t(\varepsilon) := \min \left\{ k \mid \frac{\sum_{i=t+k}^{\infty} \gamma(i)}{\sum_{i=t}^{\infty} \gamma(i)} \leq \varepsilon \right\}$$

Theorem. If $H_t(\varepsilon) \in o(t)$, then there is a limit-computable policy that is weakly asymptotically optimal in the class of all computable stochastic environments.

Limit-Computable BayesExp

Effective horizon:

$$H_t(\varepsilon) := \min \left\{ k \mid \frac{\sum_{i=t+k}^{\infty} \gamma(i)}{\sum_{i=t}^{\infty} \gamma(i)} \leq \varepsilon \right\}$$

Theorem. If $H_t(\varepsilon) \in o(t)$, then there is a limit-computable policy that is weakly asymptotically optimal in the class of all computable stochastic environments.

BayesExp: `if` $V_I^{\pi} > \varepsilon_t$ `then` execute π_I^* for $H_t(\varepsilon_t)$ steps
`else` execute π_ξ^* for 1 step

Limit-Computable BayesExp

Effective horizon:

$$H_t(\varepsilon) := \min \left\{ k \mid \frac{\sum_{i=t+k}^{\infty} \gamma(i)}{\sum_{i=t}^{\infty} \gamma(i)} \leq \varepsilon \right\}$$

Theorem. If $H_t(\varepsilon) \in o(t)$, then there is a limit-computable policy that is weakly asymptotically optimal in the class of all computable stochastic environments.

BayesExp: `if` $V_I^\pi > \varepsilon_t$ `then` execute $\pi_I^{\varepsilon_t}$ for $H_t(\varepsilon_t)$ steps
`else` execute $\pi_\xi^{\varepsilon_t}$ for 1 step

Outline

References I

Hutter, M. (2005).
Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability.
Springer.

Lattimore, T. (2013).
Theory of General Reinforcement Learning.
PhD thesis, Australian National University.

Lattimore, T. and Hutter, M. (2011).
Asymptotically optimal agents.
In *Algorithmic Learning Theory*, pages 368–382. Springer.

Leike, J. and Hutter, M. (2015a).
Bad universal priors and notions of optimality.
In *Conference on Learning Theory*, pages 1244–1259.

Leike, J. and Hutter, M. (2015b).
On the computability of AIXI.
In *Uncertainty in Artificial Intelligence*, pages 464–473.

References II

Orseau, L. (2013).
Asymptotic non-learnability of universal agents with computable horizon functions.
Theoretical Computer Science, 473:149–156.

Orseau, L. (2014).
Universal knowledge-seeking agents.
Theoretical Computer Science, 519:127–139.

Orseau, L., Lattimore, T., and Hutter, M. (2013).
Universal knowledge-seeking agents for stochastic environments.
In *Algorithmic Learning Theory*, pages 158–172. Springer.