Explicit Legg-Hutter intelligence calculations which suggest non-Archimedean intelligence

Samuel Allen Alexander & Arthur Paul Pedersen

2024

Reinforcement learning

In *reinforcement learning* (RL), an agent interacts with an environment.

1. The agent takes an *action* (trying to maximize total reward).
2. The environment responds with an *observation* and a numerical *reward*.
3. Goto 1.

Thus, RL models what Fabrizia called *conditioned learning*. The observations in RL can be thought of as what Philipp called *states*. The environment could even involve human interaction, as in Antonio's talk.

A probability paradox: The Garden of Eden

- An agent must spend eternity in the Garden of Eden.
- If it never takes the forbidden action, total reward $= 1$.
- If it ever takes the forbidden action, total reward drops to 0 (no way to recover).
- Agent A, every turn, randomly takes the forbidden action with probability 0.0001%.
- Agent B, every turn, randomly takes the forbidden action with probability 99.999%.

Paradox: Agents A and B both have total expected reward 0.

Solution using non-Archimedean probability

The Garden-of-Eden probability paradox can be solved by using nonstandard probability where infinitesimals are allowed. A and B both get positive infinitesimal total expected rewards, but A's infinitesimal total expected reward is bigger than B's.

Legg-Hutter intelligence

The Legg-Hutter intelligence of a reinforcement learning agent π is

$$\Upsilon(\pi) = \sum_{\mu} 2^{-K(\mu)} V_{\mu}^{\pi},$$

where:

- μ ranges over suitably well-behaved computable reinforcement learning environments.
- $K(\mu)$ is the Kolmogorov complexity of μ (a measure of μ's computational complexity).
- V_{μ}^{π} is the total expected reward π would get in μ.

This definition attempts to avoid all sorts of biases (such as the nationality bias from Leonardo's talk). By summing over all μ, we include rich environments with what Giovanni called *sensory richness*.

Legg-Hutter intelligence

The Legg-Hutter intelligence of a reinforcement learning agent π is

$$\Upsilon(\pi) = \sum_{\mu} 2^{-K(\mu)} V_{\mu}^{\pi},$$

where:

- μ ranges over suitably well-behaved computable reinforcement learning environments.
- $K(\mu)$ is the Kolmogorov complexity of μ (a measure of μ's computational complexity).
- V_{μ}^{π} is the total expected reward π would get in μ.

This definition attempts to avoid all sorts of biases (such as the nationality bias from Leonardo's talk). By summing over all μ, we include rich environments with what Giovanni called *sensory richness*.

But: $K(\mu)$ and thus $\Upsilon(\pi)$ depend implicitly on the background choice of a universal Turing machine.

Promoting a probability paradox to an intelligence paradox

Carefully choosing the background UTM, we can arrange that for certain agents π, including usually-well-behaved A and usually-bad-behaved B, all terms of

$$\Upsilon(\pi) = \sum_{\mu} 2^{-K(\mu)} V_{\mu}^{\pi}$$

cancel in pairs, *except where* $\mu =$ *Garden of Eden*.

- Then the *intelligence* of A and B is determined entirely by their expected Garden-of-Eden reward.
- So $\Upsilon(A) = \Upsilon(B) = 0$. The paradox appears as an intelligence paradox, not just a probability paradox.

Solving the intelligence paradox with nonstandard analysis

- We show that the intelligence measurement Garden-of-Eden paradox can be solved by measuring intelligence using the hyperreal numbers, a number system allowing for infinities and infinitesimals.
- This is evidence that $\mathbb{R}$ might be an inadequate number system for measuring intelligence of reinforcement learning agents.

Formal details: Duality

- If π is an RL agent, the *dual* agent $\overline{\pi}$ is the agent obtained from π by multiplying all rewards by -1.
- If μ is an RL environment, the *dual* environment $\overline{\mu}$ is the environment obtained from μ by multiplying all rewards by -1.
- π is *self-dual* if $\pi = \overline{\pi}$.
- μ is *self-dual* if $\mu = \overline{\mu}$.

Algebraic properties of duality

Suppose π is an RL agent and μ is an RL environment.

- $\overline{\overline{\pi}} = \pi$, $\overline{\overline{\mu}} = \mu$.
- $V_{\overline{\mu}}^{\overline{\pi}} = -V_{\mu}^{\pi}$.
- $V_{\mu}^{\overline{\pi}} = -V_{\overline{\mu}}^{\pi}$.
- If π and μ are both self-dual: $V_{\mu}^{\pi} = 0$.

Symmetric universal Turing machines

A universal Turing machine is *symmetric* if:

- For every RL environment μ, $K(\mu) = K(\overline{\mu})$.

Requiring the UTM to be symmetric solves an inherent bias in reinforcement learning. It is arbitrary whether the numbers from the environment are "rewards" (positive is good) or "costs" (negative is good). In RL, we arbitrarily choose them to be "rewards". This introduces an implicit bias. Symmetric UTMs fix this bias, in the context of Legg-Hutter intelligence measurement.

Two theorems by Alexander and Hutter

- Theorem: There is an effective procedure for turning UTMs into symmetric UTMs.
- Theorem: If the implicit background UTM is symmetric, then for any agent π, $\Upsilon(\pi) = -\Upsilon(\overline{\pi})$.
- Corollary: If the implicit background UTM is symmetric and π is self-dual, then $\Upsilon(\pi) = 0$.

(That corollary confirms what Graham said in his talk: not everything is intelligent.)

Almost-symmetric UTMs

Assume μ is an environment, $\mu \neq \overline{\mu}$.

- Definition: A UTM is *almost symmetric except at* μ if $K(\mu) \neq K(\overline{\mu})$ but $K(\nu) = K(\overline{\nu})$ for all $\nu \notin \{\mu, \overline{\mu}\}$.
- Theorem: There is an effective procedure for turning UTMs into UTMs that are almost symmetric except at μ. In fact, we can even arrange that $K(\mu) = 1$ and $K(\overline{\mu}) = 2$.
- Theorem: If the background UTM is almost symmetric except at μ, and $K(\mu) = 1$ and $K(\overline{\mu}) = 2$, and if π is a self-dual agent, then $\Upsilon(\pi) = \frac{1}{4} V_{\mu}^{\pi}$.

Garden-of-Eden probability paradox as an intelligence measurement paradox

- If A is the agent who randomly takes the forbidden action in the garden of Eden with probability 0.0001% every turn, and otherwise takes a fixed non-forbidden action, and similar for B taking the forbidden action with probability 99.999%, then A and B are self-dual (because they ignore rewards).
- Thus, if we arrange that the UTM is almost-symmetric except for the Garden of Eden environment, then $\Upsilon(A)$ and $\Upsilon(B)$ depend only on the expected total reward A and B get in the Garden of Eden. We end up with $\Upsilon(A) = \Upsilon(B)$, even though in a sense, A performs better in the only environment that counts.

Hyperreal-valued Legg-Hutter intelligence

We propose a variation on Legg-Hutter intelligence, namely:

$$\hat{\Upsilon}(\pi) = \left[n \mapsto \sum_{\mu} 2^{-K(\mu)} V^{\pi}_{\mu,n}, \right]$$

the hyperreal number represented by the function sending each $n \in \mathbb{N}$ to $\sum_{\mu} 2^{-K(\mu)} V^{\pi}_{\mu,n}$, where $V^{\pi}_{\mu,n}$ is the expected reward π would get after n turns in environment μ. This works assuming we restrict attention to environments with the property that $V^{\pi}_{\mu,n}$ never goes outside of $[-1, 1]$.

- Theorem: For any π, the difference between $\hat{\Upsilon}(\pi)$ and $\Upsilon(\pi)$ is at most infinitesimal.
- Theorem: In the context of the Garden-of-Eden paradox, $\hat{\Upsilon}(A) > \hat{\Upsilon}(B)$ (both are positive infinitesimal).

An Electoral Introduction to Hyperreal Numbers

Suppose we wish to answer True/False questions about functions $f, g : \mathbb{N} \to \mathbb{R}$. Questions like, "Is f bigger on average than g?"

An Electoral Introduction to Hyperreal Numbers

Suppose we wish to answer True/False questions about functions $f, g : \mathbb{N} \to \mathbb{R}$. Questions like, "Is f bigger on average than g?"

- The electoral approach: let each $n \in \mathbb{N}$ cast a *vote*!
- For example: If $f(25) > g(25)$ then 25 votes that f is bigger on average than g. If $f(60) = g(60)$ then 60 votes that f is not bigger on average than g.

But how can we decide an election with ∞ voters?

Majorities

What axioms should a notion of *majority* of ∞ voters satisfy?

- (Properness) $\emptyset$ is not a majority.
- (Monotonicity) If $X \subseteq Y$ and X is a majority, then Y is a majority.
- (Maximality, "Someone must win") Either X is a majority, or X^c is a majority.

A counter-intuitive axiom

If the voters vote...

- "f is bigger on average than g", and
- "g is bigger on average than h", ...

...then they had certainly better vote...

- "f is bigger on average than h"!

A counter-intuitive axiom

If the voters vote...

- "f is bigger on average than g", and
- "g is bigger on average than h", ...

...then they had certainly better vote...

- "f is bigger on average than h"!

This leads to a 4th, counter-intuitive axiom:

- ($\cap$-closure) If X and Y are majorities, then $X \cap Y$ is a majority.

Avoiding a trivial solution

One trivial way to define majorities: choose some $n \in \mathbb{N}$ and declare n a dictator. Declare that whoever n votes for, wins. Let's rule out this trivial solution.

- (Non-dictatorialness) If $|X| = 1$ then X is not a majority.

Ultrafilters

- Definition: A set of subsets of $\mathbb{N}$ (thought of as *majorities*) is an *ultrafilter* if it satisfies Properness, Monotonicity, Maximality, and $\cap$-closure.
- Definition: An ultrafilter is *free* if it also satisfies Non-dictatorialness.
- Theorem: Free ultrafilters exist.

Deciding elections using ultrafilters

Fix a free ultrafilter $\mathcal{U}$ (call its elements *majorities*). If the natural numbers vote in an election between candidates C_1 and C_2, declare:

- C_1 wins if $\{n \in \mathbb{N} : n \text{ votes for } C_1\}$ is a majority.
- C_2 wins if $\{n \in \mathbb{N} : n \text{ votes for } C_2\}$ is a majority.

By the Maximality axiom, there is a winner. By $\cap$-closure and Properness, there is at most one winner.

The hyperreals

- For all $f, g : \mathbb{N} \to \mathbb{R}$, declare $f \sim g$ if and only if a majority votes that $f(n) = g(n)$. In other words, $f \sim g$ iff $\{n \in \mathbb{N} : f(n) = g(n)\}$ is a majority.
- Lemma: $\sim$ is an equivalence relation. For each $f : \mathbb{N} \to \mathbb{R}$, let $[f]$ be f's equivalence class mod $\sim$. These equivalence classes are called *hyperreal numbers*.
- Definition: For $f, g : \mathbb{N} \to \mathbb{R}$, define $[f] + [g] = [f + g]$, $[f] \cdot [g] = [f \cdot g]$. Declare $[f] < [g]$ iff $\{n \in \mathbb{N} : f(n) < g(n)\}$ is a majority.
- Theorem: This makes the hyperreal numbers an ordered field extension of $\mathbb{R}$.

Examples of Hyperreal Numbers

- $[n \mapsto 5]$ is the hyperreal number 5.
- $[n \mapsto n]$ is an infinite hyperreal number.
- $[n \mapsto n^2]$ is a larger infinite hyperreal number.
- $[n \mapsto 1/(n+1)]$ is an infinitesimal hyperreal number.

Hyperreal-valued Legg-Hutter intelligence

We propose a variation on Legg-Hutter intelligence, namely:

$$\hat{\Upsilon}(\pi) = \left[n \mapsto \sum_{\mu} 2^{-K(\mu)} V^{\pi}_{\mu,n}, \right]$$

the hyperreal number represented by the function sending each $n \in \mathbb{N}$ to $\sum_{\mu} 2^{-K(\mu)} V^{\pi}_{\mu,n}$, where $V^{\pi}_{\mu,n}$ is the expected reward π would get after n turns in environment μ. This works assuming we restrict attention to environments with the property that $V^{\pi}_{\mu,n}$ never goes outside of $[-1, 1]$.

- Theorem: For any π, the difference between $\hat{\Upsilon}(\pi)$ and $\Upsilon(\pi)$ is at most infinitesimal.
- Theorem: In the context of the Garden-of-Eden paradox, $\hat{\Upsilon}(A) > \hat{\Upsilon}(B)$ (both are positive infinitesimal).

Conclusion

- Certain probability paradoxes become Legg-Hutter intelligence measurement paradoxes if we choose the background universal Turing machine very carefully.
- We conjecture that such paradoxes are present in Legg-Hutter intelligence even with more familiar universal Turing machines, but it is hard to exhibit them because of the intractibility of the infinite sum defining $\Upsilon(\pi)$.
- One such paradox, our Garden-of-Eden paradox, can be solved by varying Legg-Hutter intelligence to be hyperreal-valued (allowing infinities and infinitesimals).
- We submit this as evidence that infinities and infinitesimals might be inherently necessary to measure intelligence with total accuracy.