Optimistic AIXI

Peter Sunehag and Marcus Hutter

2012

Optimism


In the Twister's Path
What's behind the worst tornado season in 50 years?
Campaign 2012
The Obama engine revs up
Why Bibi bashed Barack
BY JOE KLEIN
TIME
MY KID IS A Genius
I WILL LIVE to 100
THIS INVESTMENT CAN'T LOSE
LOOK ON THE BRIGHT Side
THE SCIENCE of OPTIMISM
Hope isn't rational—so why are humans wired for it? By Tali Sharot

General Reinforcement Learning

An environment $\nu(h_t, a_t) = (o_t, r_t)$
where $h_t = a_1 o_1 r_1, \ldots, a_t o_t r_t$.

Maximize the discounted reward sum (return) $\sum_{i=t}^{\infty} r_i \gamma^{t-i}$ where $\gamma \in (0, 1)$

A policy is a function $\pi(h_t) = a_t$

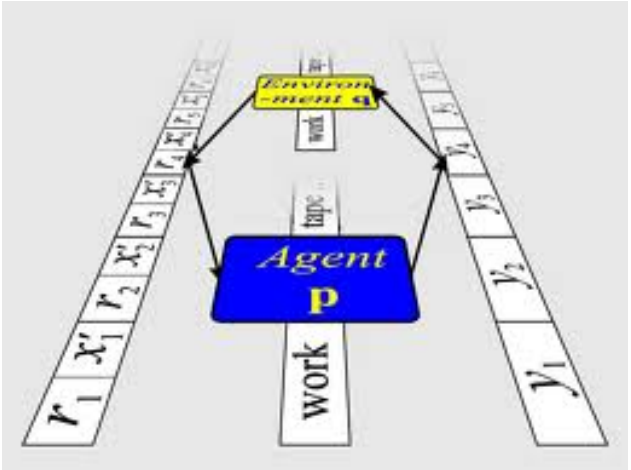
$V_\nu^\pi(h_t)$ = expected return (in ν)
achieved by following policy π after h_t

π is asymptotically optimal if for the true environment μ
$\lim_{t\to\infty}(\max_{\tilde{\pi}} V_\mu^{\tilde{\pi}}(h_t) - V_\mu^\pi(h_t)) = 0$

AIXI

The rational policies are $\arg\max_\pi V^\pi_\xi(h_t)$

- The AIXI agent defines ξ as a mixture $\sum w_\nu \nu$ over all lower semi-computable environments and $\omega_\nu > 0$ for all such ν
- The probability of generating a string as output from a Universal Turing Machine when choosing input by coin flips
- AIXI is optimal on average with respect to the prior used to define it.
- AIXI is rational (Sunehag and Hutter 2011)
- Orseau (2010): AIXI is not guaranteed asymptotic optimality
- Lattimore and Hutter (2011): no agent is but in a weaker sense it is possible though AIXI fail due to insufficient exploration



Optimism and Optimality

If $\max_\pi V^\pi_\xi(h) \geq \max_\pi V^\pi_\mu(h)$ for the true environment μ, we say that ξ is optimistic

If ξ is also dominant, i.e. if $\xi(\cdot) \geq w_\mu \mu(\cdot)$ then AIXI is guaranteed asymptotic optimality

We extend the AIXI agent by instead of picking just one ξ, we pick a compact class Ξ

- Optimistic AIXI acts according to a policy in $\arg\max_\pi \max_{\xi \in \Xi} V^\pi_\xi(h)$
- If there is at least one optimistic environment in Ξ and if all are dominant, then asymptotic optimality is achieved
- Which environment is optimistic depends on the history
- Decreased dependence on the choice of reference machine by picking a class.
- NOT the same as combining many machines into one. More explorative

Optimism and Rationality

Decision Theoretic Rationality defined as being consistent

- Morgenstern, Von Neumann
  and later Savage provide axiomatization
- Consequence:
  Preferences are deemed consistent (rational)
  if there are beliefs and tastes that explain
  the preferences as maximizing expected utility
- Suggestion:
  Break the symmetry assumption that
  implies that you do not like both sides of a bet.
- Consequence:
  Leads to our class of optimistic agents
- The information gained from the experience
  makes this reasonable in reactive environments

Contracts

We are about to observe an event.

It consists of a letter from a finite alphabet.

We are offered a bet on what it is.

Definition (Contract)

A contract is an element $x = (x_1, \ldots, x_m)$ in $\mathbb{R}^m$ and x_j is the reward received if the event (the truth) is the j:th symbol, under the assumption that the contract is accepted (see next definition).

Decision Maker, (Optimistic) Rationality

Definition (Decision Maker, Decision)

A decision maker (for some unknown environment) is a set $Z \subset \mathbb{R}^m$ which defines exactly the contracts that are acceptable (and $\tilde{Z}$ rejectable) and this we can define using a function from $\mathbb{R}^m$ to $\{\text{accepted, rejected, either}\}$. The function value is called the decision.

Definition (Rationality)

We say that a decision maker is rational if

1. Every contract $x \in \mathbb{R}^m$ is either acceptable or rejectable or both;
2. x is acceptable if and only if $-x$ is rejectable;
   we replace iff with if and rationality with optimistic rationality
3. $x, y \in Z$, $\lambda, \gamma \geq 0$ then $\lambda x + \gamma y \in Z$;
4. If $x_k \geq 0 \ \forall k$ then $x = (x_1, ..., x_m) \in Z$ while if $x_k < 0 \ \forall k$ then $x \notin Z$.

- Optimistic Rationality admits (leads to) optimistic agents

Including non-dominant environments

Consider a class of environments
Ξ that are not all assumed to be dominant

- Environments might then need to be excluded at some point
- Exclude based on threshold on likelihood ratio
- Sunehag and Hutter (2012, AusAI) prove asymptotic optimality if the true environment is in the class
- This condition replace the assumption that all the environments are dominant and this environment is automatically optimistic (since it is relative to itself)

Conclusions

- Classical rationality axioms lead to Bayesian agents
- Bayesian agents can often fail to achieve asymptotic optimality in the active reinforcement learning setting because they are insufficiently explorative
- We weaken the assumption that one does not strictly like both sides of a bet which is motivated by the possibility of gaining experience to learn from
- Leads to optimistic agents that achieve asymptotic optimality for larger classes of environments
- Optimistic AIXI is less sensitive to the initial choice of reference machine