Theoretical Computer Science 519 (2014) 127–139

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Universal knowledge-seeking agents

Laurent Orseau [a,b]

[a] AgroParisTech, UMR 518 MIA, F-75005 Paris, France
[b] INRA, UMR 518 MIA, F-75005 Paris, France

ARTICLE INFO

Keywords:
Universal artificial intelligence
AIXI
Solomonoff induction
Artificial general intelligence

ABSTRACT

Reinforcement learning (RL) agents like Hutter's universal, Pareto optimal, incomputable AIXI heavily rely on the definition of the rewards, which are necessarily given by some "teacher" to define the tasks to solve. Therefore, as is, AIXI cannot be said to be a fully autonomous agent. From the point of view of artificial general intelligence (AGI), this can be argued to be an incomplete definition of a generally intelligent agent.

Furthermore, it has recently been shown that AIXI can converge to a suboptimal behavior in certain situations, hence showing the intrinsic difficulty of RL, with its non-obvious pitfalls.

We propose a new model of intelligence, the *knowledge-seeking agent* (KSA), halfway between Solomonoff induction and AIXI, that defines a completely autonomous agent that does not require a teacher. The goal of this agent is not to maximize arbitrary rewards, but to entirely explore its world in an optimal way. A proof of strong asymptotic optimality for a class of horizon functions shows that this agent behaves according to expectation. Some implications of such an unusual agent are proposed.

1. Introduction

In 2000, Hutter proposed AIXI [1–3] the first universal and formal, though incomputable, model of an intelligent agent. It relies on the reinforcement learning framework [4] and should allow us (with computable approximations, compare [5]) to solve any practical problem as long as we are able to define the rewards.

However, it was recently proved that AIXI can in certain situations stop exploring, leading to suboptimal behavior [6]. Instead of viewing this result as a flaw of the model, given its natural definition, one can view this as a hint as to why reinforcement learning is intrinsically difficult; Furthermore the Pareto optimality of AIXI ensures that no learner (computable or not) can hope to behave better on average.

We propose a new model of a universal intelligent agent,[1] where reward signals are replaced by an internal and fully defined utility function. The goal of the agent is not to maximize the expected reward anymore, but to entirely explore the world in an optimal way. We call this kind of agent a *knowledge-seeking agent* (KSA).

Passive prediction, active learning, reinforcement learning. Solomonoff Induction is defined for passive prediction [9]: The learner does not output any action and thus cannot influence the behavior of the environment. As a consequence, learning is "easy": The predictor converges to optimal prediction in approximately $K(q)$ errors (where K is Kolmogorov's complexity [10], and q is a program that generates the sequence to predict).

E-mail address: laurent.orseau@agroparistech.fr.
URL: http://www.agroparistech.fr/mia/orseau.

[1] This paper is an extended version of [7,8].

0304-3975/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.tcs.2013.09.025

AIXI is defined in the active setting, where the environment may take into account the actions of the agent. Learning is then more difficult, because the environment may adapt its behavior to that of the agent. AIXI is actually defined in the even more difficult reinforcement learning setting, where there are rewards to maximize.

Knowledge-seeking agents are halfway between these two settings: They are active learning agents, but do not use external rewards to guide their behavior. We will show that this allows for a convergence proof in this active setting that was not possible in the RL setting.

Artificial general intelligence. Hutter described AIXI as a suitable model for universal intelligence, in the sense that this agent should be able to solve any computable problem we might give it. But regarding Artificial General Intelligence (AGI), such an agent is not fully autonomous. Indeed, in the RL framework, the rewards are given by the environment to the agent. But in our real world, the rewards are not yet defined. Therefore, if we build an RL robot, we will need to specify entirely what rewards the robot should receive, in order to define precisely what we want it to achieve. We, the designers, would then have the role of a teacher. AIXI can be viewed as a "servant" AGI agent, which must serve its teacher, who controls the rewards, whereas a KSA could be viewed as a "free" AGI agent, depending on no one.

Furthermore, we ought to be extremely careful about how we define the rewards and how they are given to a supposedly vastly intelligent RL agent like AIXI (or an approximation thereof). For example, how will the agent behave if it is not switched off when its task is done? Will it want to undo it in order to do it again? Can it be switched off, and will it resist being switched off, since this would prevent it from receiving further rewards? Or should we reward it for being switched off? If so, how can we prevent it from switching itself off to get rewards? Another related concern is whether it would try to bypass human control in order to give rewards to itself. RL agents tend to find shortcuts so as to make the minimum effort to receive rewards. This should be especially true for very intelligent agents [11,12]. We will then need to be careful that such unexpected shortcuts do not exist, which may not be a trivial matter. Hutter writes [2]:

> Sufficiently intelligent agents may increase their rewards by psychologically manipulating their human "teachers", or by threatening them.

For example, if humans use a button to control the rewards of the agent, the latter should *by all means* try to acquire the control of this button, which may lead to undesirable situations.

Hopefully all these problems have solutions, but this shows that defining rewards for a real-world RL AGI is not as simple as one may think at first.

Would a KSA be useful? A knowledge-seeking agent would not depend on any external intelligent entity, and would be fully autonomous. One drawback would be that we, humans, would have more difficulties to make it solve our specific problems, as it would have its own drives. It may still be possible to use pieces of knowledge as rewards, at least up to some point. One other possibility would be to show the agent that it would itself gain knowledge if it helped us with some particular problem. Temporarily switching off the agent could be used as a punishment, since during this time the agent cannot explore the world — although, as for AIXI, it may resist being switched off. The agent could also be biased by showing only an adequate part of the world, either real or simulated.

However, we believe it would not be the right way to use a KSA. In fact the latter may be more useful when on its own rather than directed like an RL agent with narrow goals.[2] Indeed, a KSA would need to be creative to acquire as much information about the world as possible, creating its own tools, designing its own experiments, etc. For example, it may try to come up with its own unified theory of physics, and may invent new mathematical tools. Humans could gain a lot of knowledge by working along with it, instead of directing it. It may create a lot of usable, novel byproducts in the process. Such an agent could even be viewed as the optimal scientist; Arguably, knowledge-seeking could also be seen as the core of intelligence. AGI knowledge-seeking agents would therefore be perfect complements of AGI RL agents.

The remainder of the paper is as follows. After some notation, we define a first knowledge-seeking agent, Square-KSA, and we prove convergence properties, showing that it behaves according to expectation. The second agent, Shannon-KSA, based on Shannon entropy, is then introduced, and some of its properties are exhibited, in particular a convergence proof is given, but in a more restricted fashion than for Square-KSA. We finally conclude with some remarks.

2. Notation

A string $s_1 s_2 \dots s_n \in S^n$ is a succession of $s_t \in S$ for some finite alphabet S, ordered by t. We write $s_{n:m} = s_n s_{n+1} \dots s_m$, and also $s_{<t} = s_{1:t-1}$.

At each new step t, the agent outputs an action $y_t \in \mathcal{Y}$ and the environment outputs $x_t \in \mathcal{X}$ for finite alphabets $\mathcal{Y}$ and $\mathcal{X}$ (e.g., the Boolean alphabet $\mathcal{B}$), then the next step $t+1$ begins. The *interaction pair* is written $yx_t \equiv y_t x_t$. The complete

[2] More broad goals could also be defined, but to define a *universal* goal we would need *automatic*, internal rewards, as knowledge-seeking can be thought of. Also compare [13].