Computable AIXI

MIRIxCanberra

6–7 September 2014

Abstract

This workshop is on *computable versions of AIXI*. In particular, we investigate a computable ξ, obtained by replacing Kolmogorov complexity with Kt complexity. The ultimate goal is to bring self-reflection to the AIXI formalism.

1 Introduction

AIXI considers only computable hypotheses but is itself incomputable, so cannot consider itself part of the environment. We expect that AIXI can learn and reason about parts of itself, and in this sense that it can do some kind of self-reflection. However, the AIXI model is not natural for reasoning about the self-reflection that AIXI might do. Because of this, we desire a different version of AIXI that allows self-reflection explicitly. A computable version of AIXI might be progress in this direction, since it does not exclude itself from the hypothesis space *in principle*.

In this workshop, we want to find reasonable weights w_ν in the universal semimeasure ξ that make it computable. We do this by replacing Kolmogorov complexity with the complexity measure Kt:

$$Kt(x \mid y) := \min\{\ell(p) + \log t \mid p(y) = x \text{ halting in the } t\text{-th time step}\}.$$

Choosing the weights $w_\nu = 2^{-Kt(\nu)}$, we get the semimeasure

$$\xi^{Kt}(x \mid y) := \sum_{\nu \in \mathcal{M}} w_\nu \nu(x \mid y).$$

Using ξ^{Kt} instead of ξ yields a model that we call *AIXI-Kt*.

2 Dates & Location

Dates:	Sat 6 Sep and Sun 7 Sep 2014, starting 10:00AM.
Location:	N101 in CSIT
Note:	The workshop runs all day and meals will be provided.

3 Research Questions

The following is a tentative list of research questions.

- Is ξ^{Kt} a universal semimeasure?
- Does AIXI-Kt satisfy some of the optimality properties of AIXI: Pareto optimality, balanced Pareto optimality, self-optimizing?
- What environments does AIXI succeed on but AIXI-Kt does not?
- What is the time and space complexity of AIXI-Kt?
- What if we replace $\log t$ with some other function of t?
- Are other AIXI-Kt agents in AIXI-Kt's hypothesis space?
- Can we find other computable weights w_ν that satisfy more of the desirable properties mentioned in the questions above?

4 Background Readings

In descending order of relevance:

- Introduction of Kt and Universal Search [**?**, Ch. 7.5].
- Optimality of AIXI [**?**, Ch. 5.4].
- Resource bounded complexity in general [**?**, Ch. 7.1].
- A related approach for Solomonoff Induction is the *Speed Prior* [**?**].
- Another computable approximation to AIXI is AIXItl [**?**, Ch. 7.2].