

UAI Tutorial Proposal for AGI 2013

(by Marcus Hutter, Australian National University, <http://www.hutter1.net/>)

1 Overview

Tentative title: One Decade of Universal Artificial Intelligence

Intended length: 3 hours (two sessions)

Justification: UAI is a difficult subject, and recent developments are only comprehensible after having introduced the basic background for/of UAI. So I expect Session 1 to cover background and introduction to UAI, and Session 2 to cover approximations, applications, alternatives, outlook, discussion.

Intended audience: Researchers interested in knowing about or building AGIs from first principles, and others interested in the formal foundations of intelligence.

2 Background, goal, and outline of the tutorial

The first decade of this century has seen the nascency of the first mathematical theory of general artificial intelligence. This theory of Universal Artificial Intelligence (UAI), supported by governmental grants totalling over 1 million dollar, has made significant contributions to many theoretical, philosophical, and practical AI questions.

In a series of papers culminating in book (Hutter, 2005), an exciting sound and complete mathematical model for a super intelligent agent (AIXI) has been developed and rigorously analyzed.

While nowadays most AI researchers avoid discussing intelligence, the award-winning PhD thesis (Legg, 2008) provided the philosophical embedding and investigated the UAI-based universal measure of rational intelligence, which is formal, objective and non-anthropocentric.

Recently, effective approximations of AIXI have been derived and experimentally investigated in JAIR paper (Veness et al. 2011). This practical breakthrough has resulted in some impressive applications, finally muting earlier critique that UAI is only a theory. For the first time, without providing any domain knowledge, the same agent is able to self-adapt to a diverse range of interactive environments. For instance, AIXI is able to *learn* from scratch to play TicTacToe, Pacman, Kuhn Poker, and other games by trial and error, without even providing the rules of the games.

The Kurzweil award winning AGI papers (Hutter 2009a,b) have introduced an alternative approach designed to be more easily scalable in practice. The basic idea is to automatically learn how to reduce intractable general real-world reinforcement

learning to tractable (structured) Markov decision processes which only model the reward-relevant aspects of the problem. First empirical evaluations (Phuong et al. 2011–2012) have demonstrated that this approach is competitive with current AIXI approximations.

These achievements give new hope that the grand goal of Artificial General Intelligence is not elusive.

This tutorial provides an informal overview of UAI in context. It attempts to gently introduce a very theoretical, formal, and mathematical subject, and discusses philosophical and technical ingredients, traits of intelligence, underlying assumptions, axiomatic characterization (if time), relation to other work, differences to human intelligence, some social questions, and the past and future of UAI.

3 Slides – Sources

Slides will be drawn and adapted from the following resources, and new slides with recent developments and the AGI-audience in mind will be created:

About 360 (=30h) course slides on UAI (Sec.8&9) with all background (Sec.2-7):
<http://www.hutter1.net/ai/suaibook.pdf>

Recent applications of Feature Reinforcement Learning:
<http://www.hutter1.net/publ/sctmrl.pdf>

4 Mini-Bio

Marcus Hutter is Professor in the Research School of Computer Science (RSCS) at the Australian National University in Canberra, Australia. He received his PhD and BSc in physics from the LMU in Munich and a Habilitation, MSc, and BSc in informatics from the TU Munich. Since 2000, his research at IDSIA and now ANU is centered around the information-theoretic foundations of inductive reasoning and reinforcement learning, which has resulted in 100+ publications and several awards. His book “Universal Artificial Intelligence” (Springer, EATCS, 2005) develops the first sound and complete *theory* of AI. He also runs the Human Knowledge Compression Contest (50'000€ H-prize).

See <http://www.hutter1.net/official/vitae.pdf> for his detailed hyper-linked curriculum vitae.