
COMP4620/8620: ADVANCED TOPICS IN AI FOUNDATIONS OF ARTIFICIAL INTELLIGENCE

Marcus Hutter

Australian National University
Canberra, ACT, 0200, Australia
<http://www.hutter1.net/>



ANU

Abstract: Motivation

The dream of creating artificial devices that reach or outperform human intelligence is an old one, however a computationally efficient theory of true intelligence has not been found yet, despite considerable efforts in the last 50 years. Nowadays most research is more modest, focussing on solving more narrow, specific problems, associated with only some aspects of intelligence, like playing chess or natural language translation, either as a goal in itself or as a bottom-up approach. The dual, top down approach, is to find a mathematical (not computational) definition of general intelligence. Note that the AI problem remains non-trivial even when ignoring computational aspects.

Abstract: Contents

In this course we will develop such an elegant mathematical parameter-free theory of an optimal reinforcement learning agent embedded in an arbitrary unknown environment that possesses essentially all aspects of rational intelligence. Most of the course is devoted to giving an introduction to the key ingredients of this theory, which are important subjects in their own right: Occam's razor; Turing machines; Kolmogorov complexity; probability theory; Solomonoff induction; Bayesian sequence prediction; minimum description length principle; agents; sequential decision theory; adaptive control theory; reinforcement learning; Levin search and extensions.

Background and Context

- Organizational
- Artificial General Intelligence
- Natural and Artificial Approaches
- On Elegant Theories of
- What is (Artificial) Intelligence?
- What is Universal Artificial Intelligence?
- Relevant Research Fields
- Relation between ML & RL & (U)AI
- Course Highlights

Organizational – ANU Course COMP4620/8620

- **Lecturer:** Marcus Hutter, **Assistant:** Sultan Javed Majeed
- **When:** Semester 2, 2017. Lecture/Tutorials/Labs:
Generic timetable: <http://timetabling.anu.edu.au/sws2017/>
Detailed Schedule: See course homepage
- **Where:** Australian National University
- **Register** with ISIS or Wattle or Admin or Lecturer.
- **Course is based on:** book “Universal AI” (2005) by M.H.
- **Literature:** See course homepage
- **Course Homepage:** More/all information available at
<http://cs.anu.edu.au/courses/COMP4620/>

Artificial General Intelligence

What is (not) the goal of AGI research?

- Is: Build general-purpose **Super-Intelligences**.
- Not: Create AI software solving specific problems.
- Might ignite a technological **Singularity**.



What is (Artificial) Intelligence?

What are we really doing and aiming at?

- Is it to build systems by trial&error, and if they do something we think is smarter than previous systems, call it success?
- Is it to try to mimic the behavior of biological organisms?

We need (and have!) theories which
can guide our search for intelligent algorithms.

“Natural” Approaches

copy and improve (human) nature



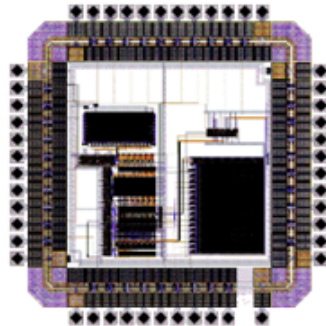
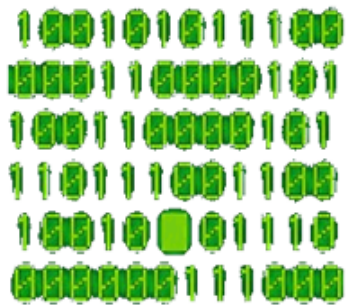
Biological Approaches to Super-Intelligence

- Brain Scan & Simulation
- Genetic Enhancement
- Brain Augmentation

Not the topic of this course

“Artificial” Approaches

Design from first principles. At best inspired by nature.



Artificial Intelligent Systems:

- Logic/language based: expert/reasoning/proving/cognitive systems.
- Economics inspired: utility, sequential decisions, game theory.
- Cybernetics: adaptive dynamic control.
- Machine Learning: reinforcement learning.
- Information processing: data compression $\approx$ intelligence.

Separately too limited for AGI, but jointly very powerful.

Topic of this course: Foundations of “artificial” approaches to AGI

There is an Elegant Theory of ...

Cellular Automata $\Rightarrow$... Computing

Iterative maps $\Rightarrow$... Chaos and Order

QED $\Rightarrow$... Chemistry

Super-Strings $\Rightarrow$... the Universe

Universal AI $\Rightarrow$... **Super Intelligence**

What is (Artificial) Intelligence?

Intelligence can have many faces $\Rightarrow$ formal definition difficult

- reasoning
- creativity
- association
- generalization
- pattern recognition
- problem solving
- memorization
- planning
- achieving goals
- learning
- optimization
- self-preservation
- vision
- language processing
- motor skills
- classification
- induction
- deduction
- ...

What is AI?	Thinking	Acting
humanly	Cognitive Science	Turing test, Behaviorism
rationally	Laws Thought	Doing the Right Thing

Collection of 70+ Defs of Intelligence

[http://www.vetta.org/
definitions-of-intelligence/](http://www.vetta.org/definitions-of-intelligence/)

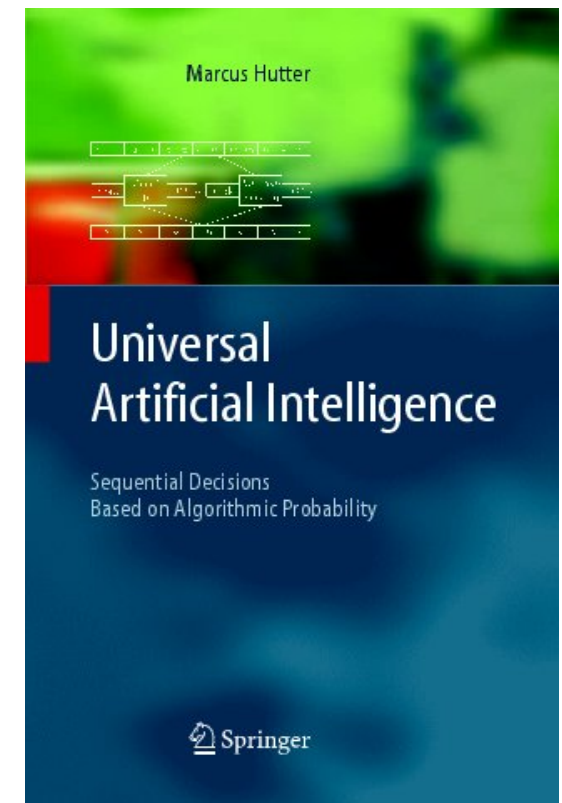
Real world is nasty: partially unobservable, uncertain, unknown, non-ergodic, reactive, vast, but luckily structured, ...

What is Universal Artificial Intelligence?

- Sequential **Decision Theory** solves the problem of rational agents in uncertain worlds if the environmental probability distribution is *known*.
- Solomonoff's theory of **Universal Induction** solves the problem of sequence prediction for *unknown* prior distribution.
- Combining both ideas one arrives at

A Unified View of Artificial Intelligence

$$\begin{array}{rcl}
 & = & \\
 \text{Decision Theory} & = & \text{Probability} + \text{Utility Theory} \\
 + & & + \\
 \text{Universal Induction} & = & \text{Ockham} + \text{Bayes} + \text{Turing}
 \end{array}$$



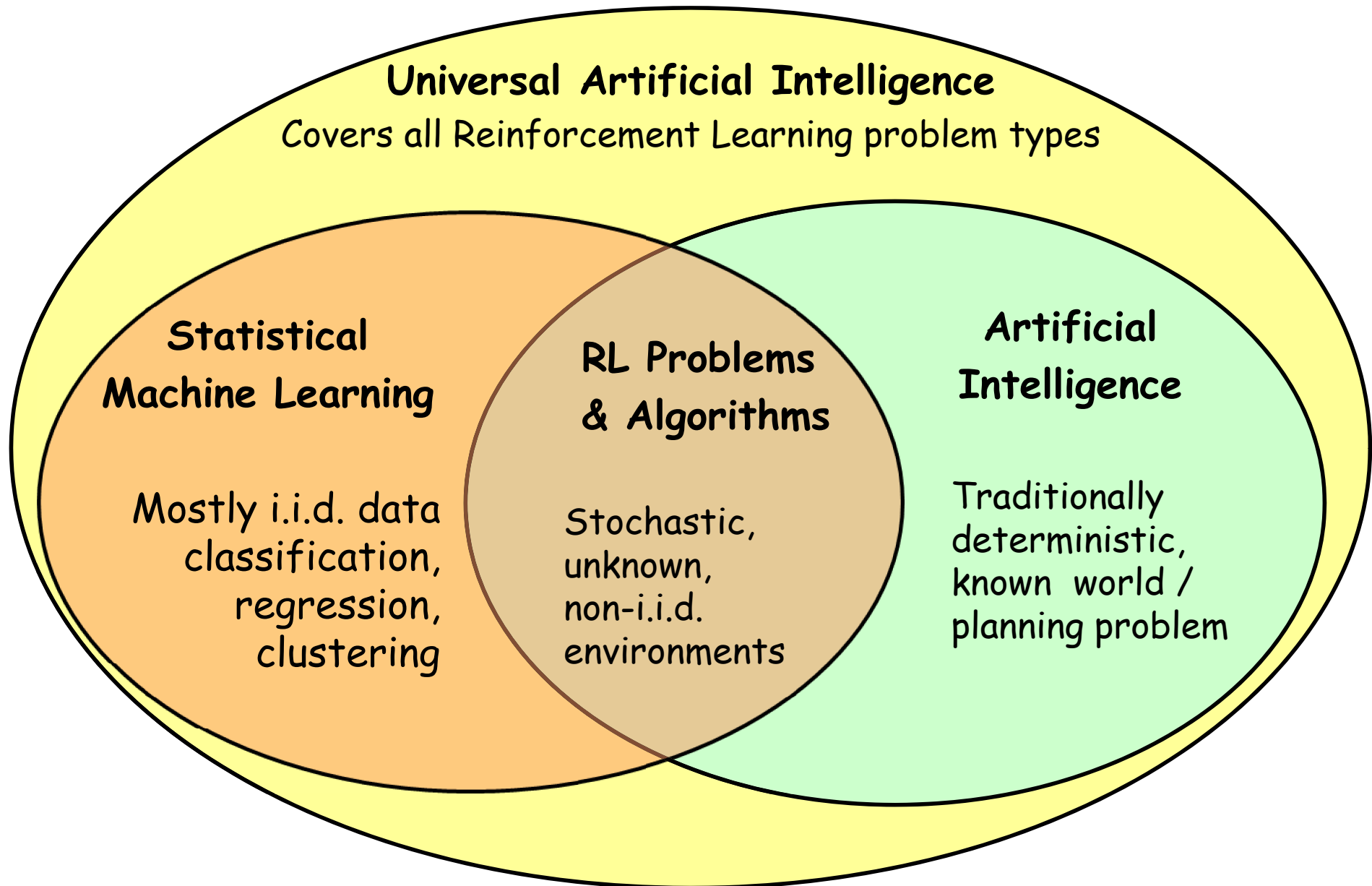
Group project: Implement a Universal Agent able to learn by itself to play TicTacToe/Pacman/Poker/... www.youtube.com/watch?v=yfsMHtmGDKE

Relevant Research Fields

(Universal) Artificial Intelligence has interconnections with
(draws from and contributes to) many research fields:

- computer science (artificial intelligence, machine learning),
- engineering (information theory, adaptive control),
- economics (rational agents, game theory),
- mathematics (statistics, probability),
- psychology (behaviorism, motivation, incentives),
- philosophy (reasoning, induction, knowledge).

Relation between ML & RL & (U)AI



Course Highlights

- Formal definition of (general rational) Intelligence.
- Optimal rational agent for arbitrary problems.
- Philosophical, mathematical, and computational background.
- Some approximations, implementations, and applications.
(learning TicTacToe, PacMan, simplified Poker from scratch)
- State-of-the-art artificial general intelligence.

Table of Contents

1. A SHORT TOUR THROUGH THE COURSE
2. INFORMATION THEORY & KOLMOGOROV COMPLEXITY
3. BAYESIAN PROBABILITY THEORY
4. ALGORITHMIC PROBABILITY & UNIVERSAL INDUCTION
5. MINIMUM DESCRIPTION LENGTH
6. THE UNIVERSAL SIMILARITY METRIC
7. BAYESIAN SEQUENCE PREDICTION
8. UNIVERSAL RATIONAL AGENTS
9. THEORY OF RATIONAL AGENTS
10. APPROXIMATIONS & APPLICATIONS
11. DISCUSSION