Universal Intelligence
According to Marcus Hutter's Definition

http://www.vetta.org/documents/UniversalIntelligence.pdf

Universal Intelligence:
A Definition of Machine Intelligence

Shane Legg

IDSIA, Galleria 2, Manno-Lugano CH-6928, Switzerland

shane@vetta.org www.vetta.org/shane

Marcus Hutter

RSISE @ ANU and SML @ NICTA, Canberra, ACT, 0200, Australia

marcus@hutter1.net www.hutter1.net

December 2007

Slides prepared by Todor Arnaudov in 2010/2011 http://research.twenkid.com

Contents

1. Human intelligence
2. Definition of machine intelligence
3. Definition of intelligence test for a machine
4. Discussion and conclusion

These slides follow the paper „Universal Intellgence“ by Marcus Hutter and Shane Legg, with a few minor notes and examples by Todor Arnaudov, who has taught it to undergraduate students as a part of two one-trimester AGI courses in *Plovdiv University, Bulgaria, Faculty of Mathematics and Informatics:*
- *Artificial General Intelligence (UAI/AGI), Spring 2010*
- *Mathematical Theory of Intelligence, Winter 2011*

About the Original Paper

M. Hutter and Sh. Legg:

The genesis of this work lies in Hutter's universal optimal learning agent, AIXI, described in 2, 12, 60 and 300 pages in [Hut01b, Hut01a, Hut05, Hut07b], respectively. In this work, an order relation for intelligent agents is presented, with respect to which the provably optimal AIXI agent is maximal.

The universal intelligence measure presented here is a derivative of this order relation. A short description of the universal intelligence measure appeared in [LH05], from which two articles followed in the popular scientific press [GR05, Fi´e05]. An 8 page paper on universal intelligence appeared in [LH06b], followed by an updated poster presentation [LH06a]. In the current paper we explore universal intelligence in much greater detail, in particular the way in which it relates to mainstream views on human intelligence and other proposed definitions of machine intelligence.

Who's smarter?

- Rex the dog, Nora the cat or Coco the parrot?
- The best in class („the nerd“) or the one who's „the life and soul“ of the parties?
- Hristo Stoichkov the striker or the academician Blagovest Sendov?
- The same person at 1, 3, 12, 25, 50, 70 years?
- Deep Blue – the chess computer or Mickey the mouse?
- Asimo the robot, or Aibo the dog?

Types of intelligence

- „Fluid“ and „Crystallized“
- Spatial
- Lingual
- Emotional (EQ)
- Mathematical
- Musical
- Muscle/Sports
- … It seems impossible to measure anything with just one measure? However...

General abstract definition of intelligence is ...

It's required in order to construct seed AI – a „fetus“ of universal self-improving intelligence.

*One functional biological suspect, having this properties are the so called neocortical mini-columns in the brains of mammals, including humans. See more in the lectures on „Brain Architecture“, „Jeff Hawkins' Hierarchical Temporal Memory“ and Boris Kazachenko's „Cognitive Focus – Specialists vs. Generalists“

Intelligence tests

- Francis Galton – reaction times.
- Binet's Test, 1905 – French, for children, 30 questions, different types, incremental difficulty: naming parts of the body, comparison, counting of coins, remembering digits and definitions of words. Has well predicted future academic susccess.
- Stanford-Binet – adopted for the USA with US military recruits tests Army Alpha & Army Beta.
- David Wechsler - 1950-s – non-verbal questions added; specific tests for different age-groups
- Progressive Matrices Test – shapes; visual (e.g. Mensa tests)

IQ

- Statistical Value – what share of people in given social group, age group or world-wide general population have higher or lower result according given test metrics.
- Impressively Stable.
- Gaussian distribution, 100 – average for given „mental age“ or a population. 10-year-old having skills of a 12-year-old one has IQ = 120.
- Mental age is discredited nowadays.

Educational Tests*

- Explicit or Implicit
- Education in kindergarten, teachers' books; games
- Detailed educational minimum for children before entering school
- Complexity of the subjects that student/subject is capable to deal with

Animal's intelligence

- Their sensory organs are developed to different extents, compared to humans *(e.g. Sense of smell is leading for dogs, while human leading one is vision)*
- Different tests for differently intelligent animals.
- Simpler – short-term and long-term memory, conditioning – association between stimuli; understanding simple patterns and prediction; counting and communication
- More Complex – is the animal capable to deceive enemy; to imitate; to recognize itself in a mirror? (Mirror Test)
- How can we make an animal do the test? By rewards... Behavior direction by operant conditioning/reinforcement learning.

Desirable properties of an IQ test

- To be repeatable.
- To be culturally neutral, lacking culture bias.

  (Verbal IQ test in foreign language, asking to fill gaps of missing words, which are rare)
- To give valid measures of intelligence.
- To be predictive – for example about future academic results of the subject
- To be easy to check.

Static and Dynamic Tests

- Static – all well known standard tests are static, they measure knowledge or skills for solving fixed questions and problems.
- Dynamic – measure subject's skills of learning and adapting.

Theories of Human Intelligence

- One general or many different capabilities?
- „Multiple factors“ - 7 primary mental abilities: verbal comprehension, word fluency, number facility, spatial visualisation, associative memory, perceptual speed and reasoning.
- Triarchic Mind – analytical, creative, practical intelligence
- „Different intelligences“ - linguistic, musical, logical-mathematical, spatial, bodily kinaesthetic, intra-personal and inter-personal intelligence
- **G-factor – general intelligence. All different components are statistically correlated.**
- „Fluid“ and „Crystallized“

Definitions about human intelligence

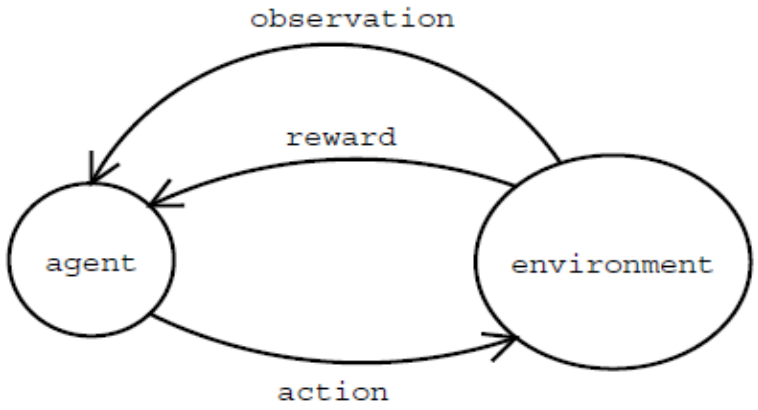
- See the paper, p.10 - 14

Definition of Machine Intelligence

- Agent, Environment, Goals.
- Communication between environment and agent:
  - agent has perceptions (inputs) from environment
  - agent acts (outputs) on environment
- How the agent could now the goal? It can be **built-in** or given as commands (**by some kind of language**).
- **Rewards – universal method, reinforcement learning.**
- Signal which tells agent whether it manages – the goal of the agent is to maximize its reward.

Definition of Machine Intelligence

- A:={left, right, up, down} – the agent sends information to the environment (interacts) – space of Actions.
- P – environment returns signals in the space of Perceptions.
- R – space of Reward [0..1]
- P:={(cold, 0.0), (warm, 1.0), (hot 0.3))}
- O – Observation
- o1r1a1o2r2a2o3r3a3o4 ...
- Observation, Reward, Action...



Functions of agent's behavior

- Π (pi) – function of the agent, whose history is input and returns next action
- **Deterministic** – always returns probability=1 for an action given the same history (it acts the same).
- **Probabilistic** (non-deterministic): returns probability between 0..1:

  Π(a3|o1r1a1o2r2) – probability to perform action a3 in the third cycle, if the history until now was o1r1a1o2r2
- Human behavior can be extrapolated or assumed as a function like this.

Function of the environment and success measure

- μ (myu) – for each k, the probability of the sequence OkRk depends on the past:

  μ(**okrk|o1r1a1o2r2a2...ak-1rk-1ak-1**)

  Measure of agent's success:

- A1 – quickly finds means to get reward of 0.9 and repeats the action, which gives it to it – **it's more successful in the beginning.**
- A2 – searches for awhile, until it founds the maximum reward 1.0 **(initially it's roaming and is less successful, however ultimately gets more successful). Period for planning future actions.**
- **„Exploration vs Exploitation“**

Formal definition of intelligence

If many hypothesis are consistent with empirical data, choose the simplest one.

- This is assumed „rational“
- Occam's Razor
- Intelligence Tests...

Dangerous „bugs“ in the environment and the agents

- From the agent's point of view, inprecise model of environment can be optimal, if the mistakes are not related to rewards.
- Occam's Razor is about the complexity of the hypothesis/theory, rather than the difficulty of carrying out a good strategy.
- Thus, in order to distinguish agents which correctly use Occam's Razor, environmental complexity should be measured, and not the difficulty of reaching the goal.

Dangerous „bugs“ in the environment and the agents

- Environment must be complex enough!
- If reward is always 1, no matter the relation between actions and observations is complex, the agent wouldn't need to search for an optimal strategy!

In human case – pleasure shouldn't come too easy and be too monotonous: addiction, drugs etc.

Kolmogorov's Complexity

- K(x) := min{ *L*(p): U(p) = x}
- p – binary string, a program
- L(p) – length of the program
- U – universal Turing machine, **reference machine**
- *K(x) – lenght of the shortest program p, which computes the string x using the reference machine*
- *There are no short programs for long random strings!*
- K is almost independent from the choice of the specific U

Real Example

- A Billion or just N of Zeros – cycle, writing N/4 times 32-bit template in the memory of a 32-bit x86 CPU.

```
INIT:  MOV edx, start_addres  //edx – start address
       MOV  eax, edx           //eax would keep...
       ADD   eax, N            //..final address
       MOV  ecx, pattern       //template
CYC:MOV  [edx], ecx            //tample goes to memory
       ADD   edx, 4            //go to next address
       CMP  edx, eax           //check for end of cycle
       JNG   CYC        //cycle again, if not end
```

Can't be done for a sequence of 1 Billion different specific numbers.

Environmental Complexity

$\mu_1, \mu_2, \mu_3, \ldots$ - different environments [virtual worlds, universes]

$\langle i \rangle$ - string, generated by a program having this level of complexity

$K(\mu_i) := K(\langle i \rangle)$ - environmental complexity

Assuming every additional bit in description decreases the probability of the environment twice.

- $2^{-K(\mu)}$ ==> Algorithmic Probability Distribution over the space of environments
- Inductive inference; the probability distribution – defining universal learning agents with provable optimal performance.

Environmental Complexity Υ

- $\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi$ - expected performance of an agent, or **universal intelligence of an agent** [ypsilon]
- **E** – space of **environments**
- **V** – **„value function“** - reward from environment
- *Internal operation of the agent is not important – any „thing“ that can output and input information and achieve goals*
- **K** (Kolmogorov's complexity) is not computable, though – can be only approximated

Υ

Types of agents depending on the environment

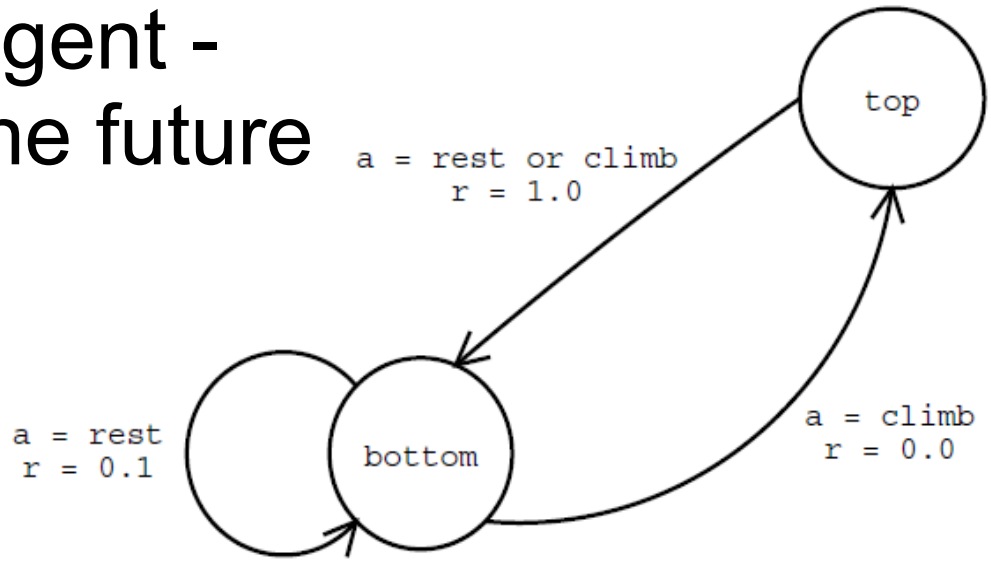
- Random $\Upsilon(\pi^{\mathtt{rand}})$
- Very specialized – Deep Blue.
- General, but simple
- Simple agent with longer history

$$\mathcal{R} = [0,1] \cap \mathbb{Q},\ \mathcal{A} = \{\mathtt{up}, \mathtt{down}\} \text{ and } \mathcal{O} = \{\varepsilon\}$$

$$\mu^{\mathtt{alt}}(o_k r_k | o_1 \ldots a_{k-1}) := \begin{cases} 1 & \text{if } a_{k-1} \neq a_{k-2} \wedge r_k = 2^{-k}, \\ 1 & \text{if } a_{k-1} = a_{k-2} \wedge r_k = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Types of agents depending on the environment

- Simple forward looking agent - computes the reward in the future



- Very intelligent agent
- Super intelligent agent – perfect theoretical agent/AIXI

$$\bar{\Upsilon} := \max_{\pi} \Upsilon(\pi) = \Upsilon\left(\pi^{AIXI}\right).$$

- Human

Bounded reward

$$V_\mu^\pi := \mathbf{E}\left(\sum_{i=1}^{\infty} r_i\right) \leq 1.$$

The sum of the rewards for the whole measured period should be bounded (limited to 1) in order to avoid distortions, because of the length of the period!

Properties of the universal intelligence

- Martin-L¨of randomness of a sequence – there's no „significant regularity“, it can't be compressed in any significant way. Similar to K.
- Zip, RAR, 7zip... OK, however, there could be some kind of complex smart pattern, which we don't suspect, but it could give big compression ratio:

  A random sequence? **1, 8, 2, 4, 5, 2, 8, 9, 6**
  However... 52896/1824 = 29 (simple number)

  „Universe and Mind 4“, T. Arnaudov 2004

- NB: A sequence „randomness“ depends on, and is according to the capabilities of the evaluator to find (recognize) regularities in the sequence.

Properties of the measure of UAI

Υ ypsilon

- Valid, Meaningful, Informative
- Wide Range $\pi^{\texttt{rand}}, \pi^{\texttt{basic}}, \pi^{\texttt{2back}}$ and $\pi^{\texttt{2forward}}$
- General
- Unbiased
- Fundamental (computation, information, complexity)
- Formal $\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi$
- Objective
- Universal – not anthropocentric
- Practical – test with approximated values

Informal definitions of intelligence

- „...the mental ability to sustain successful life.” K. Warwick, quoted in [Aso03], 2003.
- „... doing well at a broad range of tasks is an empirical definition of ‘intelligence“ - H. Masum 2002.
- Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines.”. - J. McCarthy 2004.
- Any system . . . that generates adaptive behaviour to meet goals in a range of environments can be said to be intelligent. D. Fogel, 1995.
- **...the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system’s ultimate goal. - J.S. Albus 1991 –** similar to M. Hutter's definition

Informal definitions of intelligence

- „“Intelligent systems are expected to work, and work well, in many different environments. Their property of intelligence allows them to maximize the probability of success even if full knowledge of the situation is not available. Functioning of intelligent systems cannot be considered separately from the environment and the concrete situation including the goal.“ - R. R. Gudwin '00. *Gudwin requires the system to operate in a particular way in order to be classyfied as „intelligent“, it shouldn't be just a black-box; to M. Hutter's definition, internal principles are not important.*
- „We define two perspectives on artificial system intelligence: (1) native intelligence, expressed in the specified complexity inherent in the information content of the system, and (2) performance intelligence, expressed in the successful (i.e., goal-achieving) performance of the system in a complicated environment.” - J. A. Horst'02
- „.... . the ability to solve hard problems.“ M. Minsky [Min85]
- „Achieving complex goals in complex environments.“ B. Goertzel [Goe06]

Informal definitions of intelligence

“. . .in any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity..” A. Newell и H. A. Simon [NS76]

„[An intelligent agent does what] is appropriate for its circumstances and its goal, it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation.“ - D. Poole [PMG98]

„Intelligence is the ability to use optimally limited resources – including time – to achieve goals.” R. Kurzweil [Kur00]

„Intelligence is the ability for an information processing agent to adapt to its environment with insufficient knowledge and resources.” [P. Wang 1995]

Resource limitations are important for practical reasons!

Tests for machine intelligence

- **Turing test –** the oldest, text chat between man and machine

  Many faults – biased; naive – man should imitate being a human (e.g. It could be smarter and faster than man)

  Loebner Prize – for conversation agents, „chat-bots“

- **Compression tests** - Hutter Prize – 100 MB extract from Wikipedia.
- **Linguistic complexity –** number of words used, length of the sentences, types of answers, syntactic complexity etc.

Tests for machine intelligence

- **Multiple cognitive abilities –** IBM *Joshua Blue*, и Adaptive AI *a2i2* Different tests – lingsuistic, for communication, associatiions; learning, different difficulty levels. „toddler Turing test“.
- **„Educational test“** (T. A. note) – following the detailed standards for cognitive skills for differnet ages and educational degree (check out nursery schools and kindergarten's teachers' books).
- **Competitive games –** e.g. ELO coefficient in chess.
- **Collection of psychometric tests** – application of human intelligence tests, Bringsjord & Schimanski.
  Criticism: Machine can be specialized to cope with exactly this kind of test and not being universal.

Tests for machine intelligence

- **C-Test (Complexity Test) –** prediction of the following symbol. Uses formal complexity measure of Levin Kt – similar to K, but computable, unlike K – Turing's machine should be capable to simulate it in linear time
  Criticism – static test.

Sequence Prediction Test

Complexity	Sequence	Answer
9	a, d, g, j, _ , . . .	m
12	a, a, z, c, y, e, x, _ , . . .	g
14	c, a, b, d, b, c, c, e, c, d, _ , . . .	d

Sequence Abduction Test

Complexity	Sequence	Answer
8	a, _ , a, z, a, y, a, . . .	a
10	a, x, _ , v, w, t, u, . . .	y
13	a, y, w, _ , w, u, w, u, s, . . .	y

Comparison of tests for machine intelligence

- **Valid –** capture intelligence, not anything else
- **Informative –** scalar or vector result or absolutue value, allowing comparison
- **Wide range –** cover from very low to very high
- **General –** could be applied both to a fly and to a machine learning algorithm
- **Dynamic –** take into account the ability to learn and adapt
- **Unbiased –** not be biased towards any particular culture, species etc.

Comparison of tests for machine intelligence

- **Fundamental –** not to be changed due to change in technology
- **Formal –** defined as preceisely as possible, ideally – using formal mathematics
- **Objective –** should not appeal to subjective assessments (no human judges)
- **Fully Defined –** all aspects and measures, specified
- **Universal –** not anthropocentric
- **Practical –** to be performed quickly and automatically

Comparison of tests for machine intelligence

Intelligence Test	Valid	Informative	Wide Range	General	Dynamic	Unbiased	Fundamental	Formal	Objective	Fully Defined	Universal	Practical	Test vs. Def.
Turing Test	•	·	·	·	•	·	·	·	·	•	·	•	T
Total Turing Test	•	·	·	·	•	·	·	·	·	•	·	·	T
Inverted Turing Test	•	•	·	·	•	·	·	·	·	•	·	•	T
Toddler Turing Test	•	·	·	·	•	·	·	·	·	·	·	•	T
Linguistic Complexity	•	●	•	·	·	·	·	•	•	·	•	•	T
Text Compression Test	•	●	●	•	·	•	•	●	●	●	•	●	T
Turing Ratio	•	●	●	●	?	?	?	?	?	·	?	?	T/D
Psychometric AI	●	●	•	●	?	•	·	•	•	•	·	•	T/D
Smith's Test	•	●	●	•	·	?	●	●	●	·	?	•	T/D
C-Test	•	●	●	•	·	●	●	●	●	●	●	●	T/D
Universal Intelligence	●	●	●	●	●	●	●	●	●	●	●	·	D

Table 1: In the table ● means "yes", • means "debatable", · means "no", and ? means unknown. When something is rated as unknown that is usually because the test in question is not sufficiently specified.

Test vs. Def. Finally, we note whether the proposal is more of a test, more of a definition, or something in between.

Further Reading

- Hut07b] M. Hutter. Universal algorithmic intelligence: A mathematical top→down approach. In *Artificial General Intelligence*, pages 227–290. Springer, Berlin, 2007
- [Hut05] M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2005. 300 pages, http://www.hutter1.net/ai/uaibook.htm.
- T. Arnaudov - *Faults in Turing Test and Lovelace Test. Introduction of Educational Test.* In Todor Arnaudov's Researches Blog, 17/11/2007
  http://artificial-mind.blogspot.com/2007/11/faults-in-turing-test-and-lovelace-test.html