Evaluating a Reinforcement Learning Algorithm with a General Intelligence Test

Javier Insa Cabrera[1], David L. Dowe[2], José Hernández Orallo[1]

1. *Departament de Sistemes Informàtics i Computació, Universitat Politècnica de València, Spain.*
2. *Computer Science & Software Engineering, Clayton School of I.T., Monash University, Clayton, Victoria, 3800, Australia.*

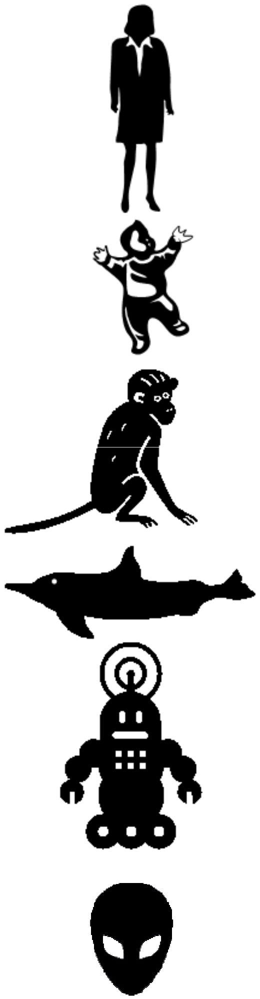
CAEPIA'2011 – San Cristóbal de La Laguna – Tenerife

7-10 noviembre 2011

Outline

- Measuring intelligence universally
- Precedents
- Test setting and administration
- Results
- Discussion
- Conclusion and future work

Measuring intelligence universally



- Can we construct a ‘universal’ intelligence test?

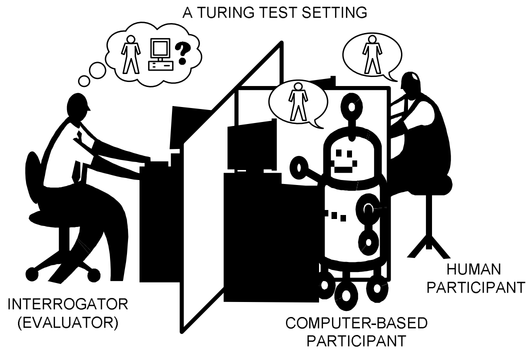
Project: **anYnt** (Anytime Universal Intelligence)

http://users.dsic.upv.es/proy/anynt/

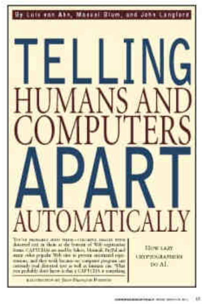
- Any kind of system (biological, non-biological, human)
- Any system now or in the future.
- Any moment in its development (child, adult).
- Any degree of intelligence.
- Any speed.
- Evaluation can be stopped at any time.

Precedents

- Imitation Game “Turing Test” (Turing 1950):
  - It is a test of *humanity*, and needs human intervention.
  - Not actually conceived to be a practical test for measuring intelligence up to and beyond human intelligence.



- CAPTCHAs (von Ahn, Blum and Langford 2002):
  - Quick and practical, but strongly biased.
  - They evaluate *specific* tasks.
  - They are not conceived to evaluate intelligence, but to tell humans and machines apart at the current state of AI technology.
  - It is widely recognised that CAPTCHAs will not work in the future (they soon become obsolete).



Precedents

- Tests based on Kolmogorov Complexity (compression-extended Turing Tests, Dowe 1997a-b, 1998) (C-test, Hernandez-Orallo 1998).
  - Look like IQ tests, but formal and well-grounded.
  - Exercises (series) are not arbitrarily chosen.
  - They are drawn and constructed from a universal distribution, by setting several 'levels' for *k*:

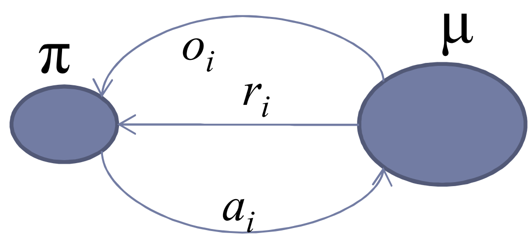
$$k = 9 \quad : \text{a, d, g, j, } \ldots \qquad \text{Answer : m}$$
$$k = 12 \ : \text{a, a, z, c, y, e, x, } \ldots \qquad \text{Answer : g}$$
$$k = 14 \ : \text{c, a, b, d, b, c, c, e, c, d, } \ldots \quad \text{Answer : d}$$

- However...
  - Some relatively simple algorithms perform well in IQ-like tests (Sanghi and Dowe 2003).
  - They are static (no planning abilities are required).

Precedents

- Universal Intelligence (Legg and Hutter 2007): an *interactive* extension to C-tests from sequences to environments.

$$\Upsilon(\pi, U) := \sum_{\mu=i}^{\infty} p_U(\mu) \cdot V_\mu^\pi = \sum_{\mu=i}^{\infty} p_U(\mu) \cdot E\left(\sum_{i=1}^{\infty} r_i^{\mu,\pi}\right)$$



= performance over a universal distribution of environments.

- Universal intelligence provides a definition which adds interaction and the notion of "planning" to the formula (so intelligence = learning + planning).
  - This makes this apparently different from an IQ (static) test.

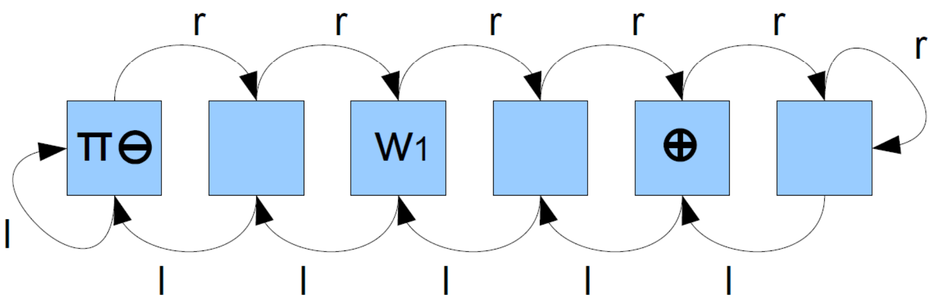
Precedents

- A definition of intelligence does not ensure an intelligence test.
- Anytime Intelligence Test (Hernandez-Orallo and Dowe 2010):
  - An interactive setting following (Legg and Hutter 2007) which addresses:
    - Issues about the difficulty of environments.
    - The definition of discriminative environments.
    - Finite samples and (practical) finite interactions.
    - Time (speed) of agents and environments.
    - Reward aggregation, convergence issues.
    - Anytime and adaptive application.
- An environment class Λ (Hernandez-Orallo 2010).

In this work we perform an implementation of the test and we evaluate a reinforcement learning algorithm with it, as a proof of concept.

Test setting and administration

- Implementation of the environment class:
  - Spaces are defined as fully connected graphs.
    - Actions are the arrows in the graphs.
    - Observations are the ‘contents’ of each edge/cell in the graph.

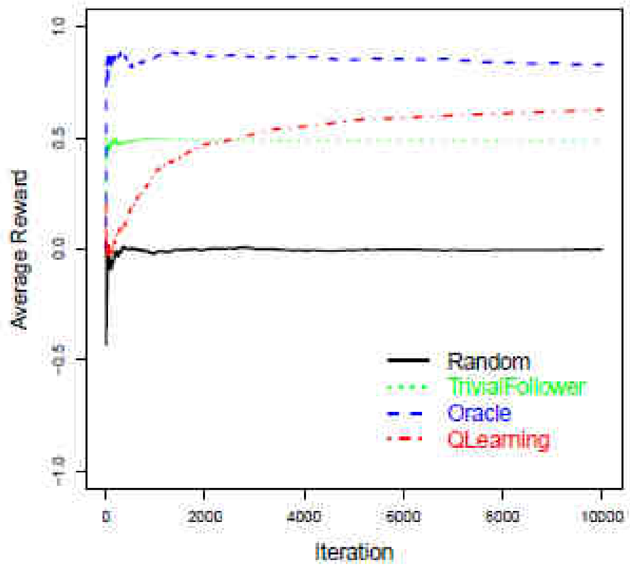


    - Agents can perform actions inside the space.
    - Rewards: Two special agents Good (⊕) and Evil (⊖), which are responsible for the rewards. Symmetric behaviour, to ensure balancedness.

Test setting and administration

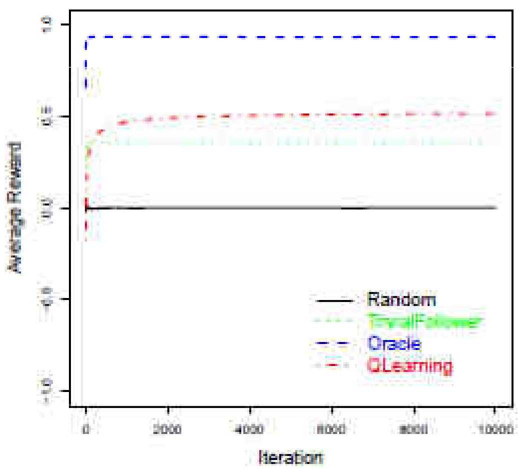
- Test with 3 different complexity levels (3,6,9 cells).
  - We randomly generated 100 environments for each complexity level with 10,000 interactions.
  - Size for the patterns of the agents Good and Evil (which provide rewards) set to 100 actions (on average).

- Evaluated Agents:
  - **Q-learning**
  - Random
  - Trivial Follower
  - Oracle

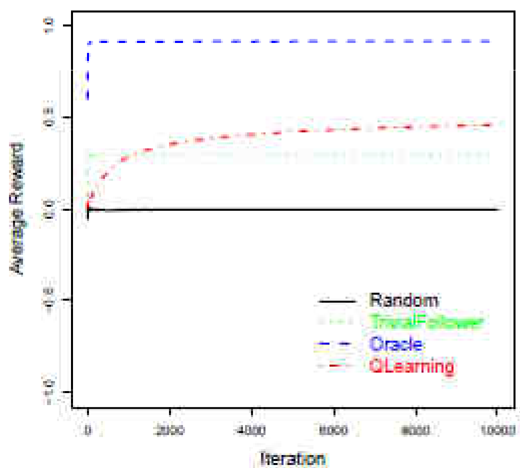


Results

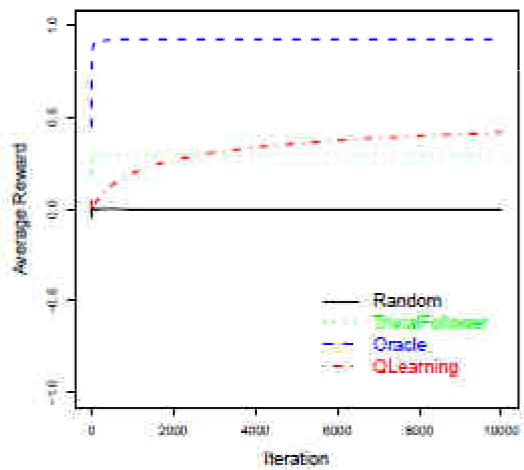
- Experiments with increasing complexity.
  - Results show that Q-learning learns slowly with increasing complexity.



3 Cells



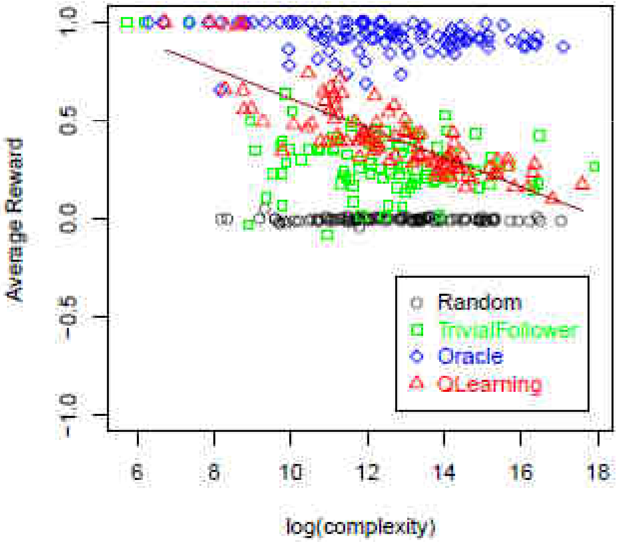
6 Cells



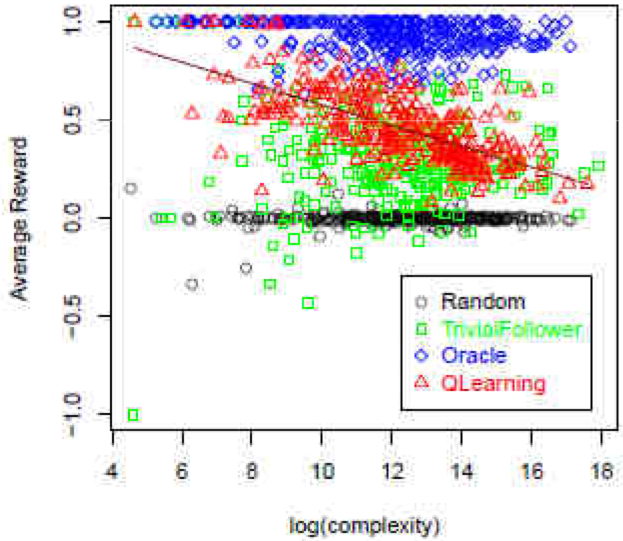
9 Cells

Results

- Analysis of the effect of complexity:
  - Complexity of environments is approximated by using (Lempel-Ziv) LZ(concat(S,P)) x |P|.



9 Cells



All environments

  - Inverse correlation with complexity (difficulty ↑, reward ↓).

Discussion

- An implementation of the Anytime Intelligence Test using the environment class Λ can be used to evaluate AI systems.
- Environment complexity is based on an approximation of Kolmogorov complexity and not on an arbitrary set of tasks or problems.
  - So it's not based on:
    - Aliasing
    - Markov property
    - Number of states
    - Dimension
    - …
- The test aims at using a Turing-complete environment generator but it could be restricted to specific problems by using proper environment classes.

Conclusion and future work

- The goal was not to analyse Q-learning, nor to designate a ‘winning’ algorithm. The goal was to show that a top-down (theory-derived) approach can work in practice.

- Future work:
  - Evaluation of other reinforcement learning algorithms and their parameters (RL-glue).
  - Progress on a new version of the implementation of the test which could be more adherent to its full specification.
    - Turing-complete environment generators.
    - Better approximations for complexity.

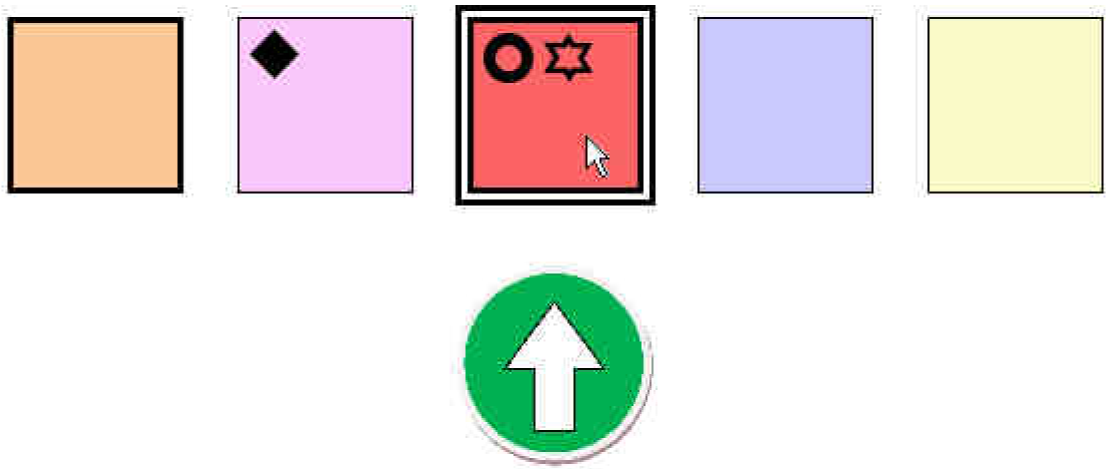
Thank you!

Some pointers:

- Project: **anYnt** (Anytime Universal Intelligence)

  http://users.dsic.upv.es/proy/anynt/

- Have fun with the test.



http://users.dsic.upv.es/proy/anynt/human1/test.html