On more realistic environment distributions for defining, evaluating and developing intelligence

José Hernández Orallo[1] , David L. Dowe[2], Sergio España-Cubillo[1], M.Victoria Hernández-Lloreda[3], Javier Insa-Cabrera[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.*
3. *Departamento de Metodología de las Ciencias del Comportamiento, Universidad Complutense de Madrid, Spain*

AGI'2011 - Mountain View, California, 3-7 August 2011

Outline

- General performance, general distribution
- Generating more social, ‘natural’ environments
- Darwin-Wallace Distribution
- Approximations
- Discussion

General Performance, General Distribution

Intelligence as performance in a wide range of tasks

- Artificial (Specific) Intelligence focusses on *specific* tasks.
  - The development of successful agents in these domains usually entails a specialised approach.
  - Problem repositories for each domain are used to evaluate these agents or algorithms (pattern recognition, machine learning, games, natural language, robotics, etc.).
  - There are very few cases in the literature where the set of problems is obtained by a *problem generator* from a specific distribution.

General Performance, General Distribution

- Artificial General Intelligence must focus on *general* tasks.
  - We can construct a general set of tasks by aggregating several problems which humans face everyday.
    - Arbitrary approach (how many of these, how many of those, ...)
    - Makes it difficult to know what “intelligence” really means.
  - But we *can* formally define a general distribution and generate tasks or environments from it.

General Performance, General Distribution

- Let us choose the most general one: a *universal distribution* over tasks or environments.

$$p_U(x) := 2^{-K_U(x)}$$

  - Where *K* is a measure of complexity (Kolmogorov complexity, or any computable approximation, Levin's *Kt*, Schmidhuber's Speed Prior, etc.)

- This approach has been explored in many ways:
  - Compression-extended Turing Tests (Dowe & Hajek 1997a-b, 1998).
  - Formal definition of intelligence, C-test (Hernandez-Orallo 1998, 2000).
  - Compression tests (Mahoney't text compression test 1999, Jim Bowery's Cprize 2005, Hutter's Prize, 2006).
  - Universal Intelligence (Legg & Hutter 2007).
  - Anytime Intelligence Tests (Hernandez-Orallo & Dowe 2010).

General Performance, General Distribution

- A universal distribution.
  - Advantages:
    - We can assign probabilities to an infinite number of tasks.
    - Universal distributions “dominate” all other possible distributions.
    - Sound results (Solomonoff’s theory of prediction, Hutter’s AIXI, etc.).
    - Simple environments frequent ⇨ Tasks easier to generate and use.
  - Disadvantages:
    - The arbitrary choice of the reference machine is still important.
      - This can be minimised by using background knowledge or using simplest UTMs (Wallace 2005, Dowe 2008a).
    - Any environment of interest (e.g. multi-agent system) has a very low probability for almost every reference machine.
      - Performance in social, natural environments, including other (intelligent) agents will not be measured.

Generating more social, ‘natural’ environments

- But intelligence is all about *social* cognition!

The Social Cognition / Cultural Intelligence Hypothesis

[Herrmann et al. 2007]

- Alternative proposals:
  - More realistic (but simplified) worlds, not using a universal distribution:
    - Social, natural, embodied environments... (e.g. AGI preschool [Goertzel 2009])
  - Choose a very particular reference machine, keeping a universal distribution:
    - Games (Hernandez-Orallo & Dowe 2010).
  - “Alter” a universal distribution:
    - Include other agents.
    - *Evolve* the distribution.

Darwin-Wallace Distribution

- We define a distribution over *multi-agent* environments (not including the agents):

$$p_E(\mu) := 2^{-K_{U_e}(\mu)}$$

- We define a distribution over agents (a "mind distribution"):

$$p_A(\pi) := 2^{-K_{U_a}(\pi)}$$

  - We assume all the agents are physically equal.
    - This is important and very different to natural evolution.
    - We only care about their "minds".

- We combine these two distributions...

Darwin-Wallace Distribution

- The probability of the *start-up* multi-agent environment σ is:

$$p_S(\sigma) = p_S(\langle \mu, \pi_1, \pi_2, ..., \pi_m \rangle) := p_E(\mu) \times \prod_{j=1}^{m} p_A(\pi^j)$$

- And now we evolve this in the following way:
  - Agent survival depends on a function *d*, related to their average rewards.
    - Dead agents are replaced by new agents.
  - The environment can be replaced by any other environment in p_E with a rate of replacement of *c*.
    - Agents do not specialise in *one* environment. They adapt to changing environments.
- The Darwin-Wallace distribution for *d*, *c* at iteration *i* is given by:

$$p_{d,c,i}(\sigma) = p_i(\langle \mu, \pi_1, \pi_2, ..., \pi_m \rangle) := p_E(\mu) \times \prod_{j=1}^{m} q_{(d,c,i)}(\pi^j)$$

  - Where *q*(*d*,*c*,*i*) is the agent probability at iteration *i*.

Darwin-Wallace Distribution

- What does this family of distributions mean?
  - It just assigns probabilities to multi-agent environments.
  - Complex agents with complex/adaptive behaviour are much more likely in this distribution, for large values of *i*.
  - The distribution is completely different for low and high values of *i*.
    - Highly social agents may be unsuccessful in environments with very simple agents, where co-operation and language are useless.
      - As a single human on an island, in the Precambrian period or on Mars.
  - Social adaptability instead of adaptation to one single environment.

Previous definitions and tests of intelligence using a universal distribution could be re-understood with a Darwin-Wallace distribution.

Approximations

- Appealing as an abstract concept.
  - Problems for using it in practice:
    - The definition is a product of other distributions, which are not necessarily independent (it would require a normalisation).
    - The distribution is uncomputable (with *K* being Kolmogorov Complexity) or clearly intractable using computable variants of *K*.
    - Some evolution “accelerators” have been ruled out (mutations, cross-over, genotype, ...).
      - We cannot wait some billion years.
  - But...

Nobody is saying that we have to wait until the agents are “naturally” created by evolution.

Approximations

- Approximation through *testing*:
  - Research-driven evolution instead of natural evolution.
    - Agents can be created artificially (by AGI researchers) but assessed in an independent way.
  - The “intelligence”/”adaptability” of agents can be assessed for different values of *i*.
    - We certify agents at lower levels of *i*, before including them in the testbed.
- This (competitive) process can foster the development of more and more (socially) intelligent systems.

Discussion

- The Darwin-Wallace distribution is not a distribution of "life forms"
  - A distribution of 'life forms' gives higher probability to bacteria and cockroaches.

- The Darwin-Wallace is a distribution of (social) "mind forms".

  - There are three features which make this distinction:
    - i) Physical traits do not matter (no body).
      - Focus is placed on behaviour.
    - ii) There is no genotype, cross-over, mutation, etc.,
      - Selection does not work for genes or species, but for individuals.
    - iii) Environments are replaced.
      - Avoids specialisation in a single environment.
      - Instead, adaptability to a wide range of environments (i.e., intelligence) is the only fitness function for selection.

Discussion

The Darwin-Wallace distribution assigns probabilities to agents depending on their success on a variety of environments with a variety of other agents.

- It relates intelligence to evolution, without abandoning the context of universal distributions.
  - This, of course, raises more questions than it answers, but...
    - It can help understand why universal distributions may be “too general” and unrealistic for worlds where intelligence has developed.
    - It can help suggest ways to link intelligence definitions with evolution, adversarial learning, competition and collaboration.

Thank you!

Some pointers:

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

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