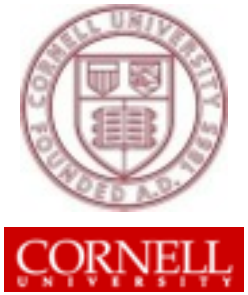

INTELLIGENCE EXPLOSION: SCIENCE OR FICTION?

Bart Selman
Cornell University

Change in Perception

2008-2009 AAAI Presidential Panel on Long-Term AI Futures

Goal: Explore societal impact of (future) AI technologies

Asilomar meeting modeled after 1975 meeting on Recombinant DNA.

General sentiment was that developments in the field required careful study but no high urgency for immediate action. (Risks warnings came from outside AI, and were felt to be too alarmist and perhaps "over-hyped.")

5 years later… Change in sentiment among a good fraction of AI researchers. Progress appears to be accelerating dramatically.

Reasons for Change

--- series of events

--- main one: machine perception is starting to work (finally!)

systems are starting to "see" and "hear"

after 50+ yrs…

--- dramatic change: lots of AI techniques (reasoning, search, reinforcement learning, planning, decision theoretic methods) were developed assuming perceptual inputs were "somehow" provided to the system. But, e.g., robots could not really see or hear anything…

(e.g. 2005 Stanley car drove around "blind", Thrun)

Now, we can use output from a perceptual system and leverage a broad range of existing AI techniques.

Our systems are finally becoming "grounded in the world," addressing the symbol-grounding problem.

Opportunities

Example: RoboBrain project (Saxena et al.)

--- goal: build a shared, large knowledge base for robots to be able to function in human environment (e.g. house) ("assistive robotics")

--- robot learns to recognize human activities from video input ("activity recognition")

--- *anticipates* human behavior for collaboration

--- uses planning and reasoning techniques to synthesize action sequences with hundreds of actions (leveraging existing technology)

robots are starting to learn and plan in new and complex behaviors in rich unconstrained environments

Factors in accelerated progress, cont.

deep learning / deep neural nets

- **success is evidence in support of the "hardware hypothesis" (Moravec)**
  - **core neural net ideas from mid 1980s**
  - **needed: several orders of magnitude increase in computational power and data**
  - **(aside: this advance was not anticipated/predicted; many AI/ML researchers had moved away from neural nets…)**

crowd-sourced human data --- machine need to understand

- **our conceptualization of the world**

engineering teams (e.g. IBM's Watson)

- **strong commercial interests (Google, Facebook, Baidu)**
- **at a scale never seen before in our field**

Next Phase

Hypothesis: Further integration of existing techniques --- perception, (deep) learning, inference, planning --- will be a game changer for AI systems.

There are still important unknowns

--- can we get at deeper semantics of natural language?

--- commonsense reasoning?

Example:

"The large ball crashed right through the table because it was made of Styrofoam."

What was made of Styrofoam? The large ball or the table?

(Oren Etzioni, Allen AI Institute)

How to involve AI research community

Although more AI researchers now see human-level AI and beyond as approaching, no consensus on what should be done about this.

Contrast with Asilomar Recombinant DNA meeting:

1) Experiments had immediate risks to researchers and population

2) Containment of experiments and other physical safety measures provided a significant step towards dealing with the risks, including a ban on certain experiments.

The AI community could reach consensus on, for example, a ban on the development of autonomous lethal drones. *But, it will be much harder to restrict underlying technologies* (e.g. face recognition --- already developed --- or autonomous systems --- needed in automated driving / trading systems).

Involving AI community, cont.

A more promising direction would be to agree on adding safeguards to AI systems to place their behavior within certain acceptable boundaries.

Friendly AI by design (Russell).

Safeguards could involve provable constraints on the system's behavior (the system verification problem; problem of characterizing new or unpredictable environments) or even separate AI modules that "watch" over a running AI system.

Many interesting research opportunities for the community.