Envisioning a Robust, Scalable Metacognitive Architecture Built on Dimensionality Reduction

Scruffy Metacognition

Jason B. Alonso¹ Kenneth C. Arnold²,³ Catherine Havasi³,²

¹Personal Robots Group
MIT Media Laboratory

²MIT Mind Machine Project

³Software Agents Group
MIT Media Laboratory

AAAI-10 Workshop on Metacognition for Robust Social Systems
July 2010
The Challenge

- Metacognitive architectures too expensive to explore iteratively
  - Programmers never count to 3.
How do we build a **scalable** framework for metacognitive architectures?

- Hand-crafted metareasoners are out.
- Use many instances of a few types of simple but powerful reasoning units.
  - Difference between reasoner and metareasoner in the inputs
  - Communicate with simple symbols, generally opaque semantics
What “reasoning units”?

What function should each component perform?

- Connectionist answer (switches, or “neurons”) theoretically satisfying to some, practically less than enlightening
- Our answer: pattern discovery and matching
  - One basic process of an intelligent system is to identify useful patterns in its input and its output
  - One symbol ⇐⇒ one pattern
    - Summarizing many inputs and outputs with fewer symbols
    - ... in essence, dimensionality reduction
We claim...

- Planning can be a pattern completion problem that leverages dimensionality reduction.
- **Metacognitive** functions, particularly metaplaning, can be built on these principles.
We are “Scruffies”

- Imprecise and loosely statistical handling of symbolic representations
We are “Scruffies”

- Imprecise and loosely statistical handling of symbolic representations

Open Mind Common Sense
We are “Scruffies”

- Imprecise and loosely statistical handling of symbolic representations

ConceptNet

Alonso, Arnold, Havasi (MIT)  Scrufty Metacognition  Metacognition 2010  8 / 21
We are “Scruffies”

- Imprecise and loosely statistical handling of symbolic representations

<table>
<thead>
<tr>
<th></th>
<th>cat</th>
<th>dog</th>
<th>airplane</th>
<th>toaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>... IsA pet</td>
<td>+6</td>
<td>+5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>... AtLocation home</td>
<td>+8</td>
<td>+2</td>
<td></td>
<td>+1</td>
</tr>
<tr>
<td>... CapableOf fly</td>
<td>−3</td>
<td>−5</td>
<td>+9</td>
<td></td>
</tr>
<tr>
<td>... MadeOf metal</td>
<td></td>
<td></td>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>fur PartOf ...</td>
<td>+6</td>
<td>+5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*AnalogySpace*, prepared
We are “Scruffies”

- Imprecise and loosely statistical handling of symbolic representations (dynamic representations)

Analysis

Alonso, Arnold, Havasi (MIT)
Sheep games

“Nexi, come take the sheep.”

Alonso, Arnold, Havasi (MIT)
Sheep games, continued

Now imagine two games...

- Both involve picking up toys and putting them somewhere
- How could Nexi know which game we’re playing?
An architecture

- Observation
- Action
- Planner Enable
- Planner Convergence
- Multiple symbols

Alonso, Arnold, Havasi (MIT)
The Reducer

- Candid, Covariance-free Incremental Principal Component Analysis (CCIPCA) (Weng et al. 2003)
- Essentially AnalogySpace
  - Dynamically-generated representations
  - Open domain
The Planner, part 1 of 2

- **Build a model** of salient patterns in observable events and behaviors
- **Generate plans** that achieve goals given this model
- **Incremental.** Learn/refine models from experience in real time
- **Scruffy.** Statistical handling of symbolic representations of the real world to draw robust conclusions
- **In practice, two approaches:**
  - Replay of natural responses to environment and teammates
  - Goal-seeking
The Planner, part 2 of 2

- Observation
- Action
- Planner Enable
- Planner Convergence

- Bus of observations
- Single action
- Reduced timeline
- Multiple symbols

EN Planner Enable
CONV Planner Convergence
Intuitive Introduction to MIDAWT

Given a multivariate waveform (timeline):

- **Detect** instances of previously-seen patterns
- **Refine** models for those patterns (or record new pattern)
- **Describe** timeline as a combination of understood patterns
- **Complete** timeline by interpolating gaps in timeline

$$\tilde{f}[t] = (\sum_i a_i \ast \tilde{g}_i)[t]$$
An architecture
The difference between cognition and metacognition is in the wiring, permitting scalable architectures.

Systems that build their own representations dynamically are more robust.

Forthcoming experiments
- Mars Escape
- Restaurant Game
- Explore/exploit
Here be dragons
A Unique MIDAWT/SP Insight

- When the astronaut player is engaged in a search activity with a bunch of boxes, the robot is **not** about to hit the elevator call button.

- Correlation not causal, but reflective of teaming behavior

- Anti-correlation not found in CBP or plan networks