GALIS
Gated Attractors Learning Instruction Sequences

Jared Sylvester
7 May 2013
Motivation

- **Gap** between neural & symbolic AI systems
  - Neural: perception, motor control, …
  - Symbolic: planning, goals, rules, …

- Neural systems are too “hard-wired”
  - Behavior is baked into architecture
  - New problems require entirely new systems
USN Mk I
Fire Control Computer
1936–1969

Jacquard Loom
1801–
Goal

- Generalizable model of cognitive control
  - Learned, not hard-wired into network structure
  - Base behavior on memory contents
- Two type of memory/learning:
  - Memory of perceptual stimuli
  - Memory of task procedures
- Biological inspiration:
  - Network of regions, recurrent attractor nets, gating, distributed representations, Hebbian learning
Attractor Net Memories

- Stored patterns are attractors
  - Form auto-associative memory
- But fixed-point attractors
  - Network gets “stuck” in attractor basin

\[ w_{ij}^t = (1 - k_D)w_{ij}^{t-1} + \frac{1}{N}a_i^ta_j^t(1 - \delta_{ij}) \]
Sequential Attractor Nets

- Dynamic thresholds
  - Increase when node’s state remains unchanged
  - Harder for node to stay in the same state


Ordered Sequential Attractors

- **Asymmetric weights**
  - Correlate activity with other nodes’ *previous activity*
    \[ v_{ij}^t = (1 - k_D)v_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^{t-1} \]
  - Network transitions between attractors in order

Adding Cognitive Control

- Modeled Running Memory Span task
  - Can match human behavioral results
  - But all control was **exogenous**
- For internal control, use **multiple networks**
  - Network of attractor networks
  - Controlled by gating
  - Learn processing of sequences
Control Mechanism

- Built around attractor networks
- Trained prior to task beginning
- Directs the model by operating gates
- Core is “Instruction Sequence Memory”
Control Mechanism

- Built around attractor networks
- Trained prior to task beginning
- Directs the model by operating gates
- Core is “Instruction Sequence Memory”
  - Stores sequence of steps to do subtasks
  - Multiple sequences stored simultaneously
  - Divided into cue & response sections
Instruction Sequence Memory

Distributed ‘cue’ pattern

Make tea
Instruction Sequence Memory

Distributed ‘cue’ pattern
- Make tea

Distributed ‘response’ patterns
- Boil water
- Steep tea bag
- Add sugar
Instruction Sequence Memory

Distributed ‘cue’ pattern

Distributed ‘response’ patterns

Make tea

Boil water

Make tea

Steep tea bag

Make tea

Add sugar
Instruction Sequence Memory

Distributed ‘cue’ pattern

Make tea ↔ Boil water

Distributed ‘response’ patterns

Make tea ↔ Steep tea bag

Make tea ↔ Add sugar

\[ w_{ij}^t = (1 - k_{CTRL})w_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^t (1 - \delta_{ij}) \]
Instruction Sequence Memory

Distributed ‘cue’ pattern  Distributed ‘response’ patterns

Make tea  Boil water

Make tea  Steep tea bag

Make tea  Add sugar

\[ v_{ij}^t = (1 - k_{CTRL})v_{ij}^{t-1} + \frac{1}{N}a_i^t a_j^{t-1} \]
**n-Back**

- Given sequence of inputs…
  …does most recent input match input \( n \) steps ago?
- Must maintain sequence in WM; make judgments

\[ n=1, 2, 3, 4, 5 \]

**GALIS model learns** \( n=1, 2, 3, 4, 5 \)

- Learns all five without knowing which it will perform
- Version determined by input patterns only

3-back

A D J A E F D K J D C A K J F F H D F

Match
No Match
No Match

(first) (last)
Controller

n Input Layer (64)  Context Module (32)

Encoder Input (96)  Encoder Hetero-associative (320)  Encoder Auto-associative (320)

Encoder Input (96)  Encoder Hetero-associative (320)  Encoder Auto-associative (320)

Cue Nodes (320)  Response Nodes (192)

INSTRUCTION SEQUENCE MEMORY (512)

Action Selection (7)  Decoder Output (6)

Output to Gates
GALIS Architecture

Visual Input

Working Memory

Compare Module

Context Module

Control Module

Memory Input Gate

Memory Training Gate

Memory Unlearning Gate

Encoder Update Gate

Output Gate

$n$ Input ("goal")

"Match" Output Node

"No-Match" Output Node
Human Comparison: Accuracy

![Bar chart showing human accuracy compared to model V and model C across different back conditions. The accuracy values range from 0.60 to 1.00.]
Human Comparison: Response Time

Model time steps per stimulus vs. Human Response Time (ms)

- $n=1$
- $n=2$
- $n=3$
Changing $n$
Visuospatial Architecture
Visuospatial Architecture
Symbolic / Sub-symbolic

- Attractor space is very high dimensional
  - Learning algorithms
  - Partial pattern matching
- Each attractor is a discrete symbol
- Gating also adds discreteness
Instruction vs. Construction

- Behavior based on memory contents not just architecture

- Can “program” a neural net
  - Now programs are hand-crafted by modeler
  - Store → improve → learn ab initio