Time-Invariant Recognition of Spatiotemporal Patterns in a Hierarchical Cortical Model with a Caudal-Rostral Persistence Gradient

Gerard J. Rinkus & John E. Lisman
Lisman Laboratory
Volen Center for Complex Systems
Brandeis University, Waltham, MA
rinkus@comcast.net

Abstract

We describe a hierarchical connectionist model of cortex that performs recognition and recall of spatiotemporal patterns (sequences). Our goal was to find an architecture invariant to presentation speed. A key feature of our model is that persistence, i.e., the length of time a unit (neuron) tends to remain active when it becomes active, increases from the bottom hierarchical level (i.e., primary sensory cortices) to higher levels (i.e., more rostral cortices). A growing body of evidence supports long persistence times for rostral neurons (Fuster & Alexander, 1971) and, more generally, a caudo-rostral persistence gradient (Uusitalo et al., 1996). In the simulations reported here, persistence doubles from one level to the next (the levels are bidirectionally connected). Thus, a set of cells that becomes active at level LJ (an LJ code) will be active during, and therefore form top-down (TD) and bottom-up (BU) associations with, two successive LJ-1 codes. The model's internal levels are organized into winner-take-all competitive modules (CMs). We provide simulation results demonstrating that the model can learn multiple complex sequences (sequences in which the same state can recur multiple times) and then recognize novel sequences that are non-uniformly time-warped and noisy versions of the previously experienced sequences. The model's robustness to time-warping results from the fact that LJ+1 codes are active during several successive LJ-1 codes. The occurrence of any of those LJ-1 codes while LJ+1 is active will result in recognition.

1: Model Overview

- Hierarchical
- Sparse distributed representations
  - Winner-take-all (WTA) competitive modules (CMs)
- Binary cell activity levels
- All binary weights except the BU weights from L1 to L2.
- Cells at higher levels have longer persistences
  - When they become active, they tend to stay active for a longer period.
- Unsupervised Hebbian learning

- Every cell in every CM receives input from all:
  - Previously active cells in neighboring levels.
  - Previously active cells in its own level.

- Activity marches up through the levels once the input is presented

2: Model Operation

On every time step and at every level:

Every cell computes:
1. the degree of match, χ, between its bottom-up (U), top-down (D), and horizontal (H) input vectors.
2. The overall degree of match, ^JG, for an entire level, J, is computed.
   - ^JG is the average of the maximal χ values in all of J's CMs.
3. Within each CM, the vector of χ values is transformed into a probability distribution (ρ values).
   - The shape of the ρ distribution depends on ^JG.
4. Within each CM, a winner is chosen on the basis of the ρ distribution.

$$\chi_t(i) = \Phi_t(i)^H \Psi_t(i)^U \Lambda_t(i)^W$$

Normalized Likelihood Vector

- Here the green circle represents the maximal normalized TD input.
- The orange 'crown' gives a visual depiction of the relative TD likelihoods over the cells of the CM.

$$\Lambda_t(i) = \frac{\lambda_t(i)}{\max\limits_{k \in CM}(\max(\lambda_t(k)),\,{}^J\Theta_J)}$$

$$\Psi_t(i) = \frac{\psi_t(i)}{\max\limits_{k \in CM}(\max(\psi_t(k)),\,{}^U\Theta_J)}$$

$$\Phi_t(i) = \frac{\phi_t(i)}{\max\limits_{k \in CM}(\max(\phi_t(k)),\,{}^H\Theta_J)}$$

χ values get turned into probability (of being chosen) values

- $\pi(z)$ is the maximal χ-value in CM z

$$\pi_t(z) = \max\limits_{i \in CM} \chi_t(i)$$

- ^JG is the overall match, at level J, between the TD, H and BU vectors.

$$G_J = \frac{\sum_z \pi_t(z)}{Q}$$

- G is implicitly a spatiotemporal similarity metric because it includes the H influence, which encodes information about the history of the sequence leading up to the current moment.
- Specifically, it is a measure of the distance between the current input sequence and the closest-matching previously-experienced sequence.

G near 1.0 leads to an expansive nonlinearity

G near 0.0 leads to an compressive nonlinearity

- The χ-values near 0 map to ξ-values at or near 1.0.
- The χ-value at 1.0 maps to a ξ-value of 4001.0

$$\xi_t(i) = \frac{4000}{1 + e^{-\cdots}} + 1$$

- All χ-values map to ξ-values at or near 1.0.

$$\xi_t(i) = \frac{1}{1 + e^{-\cdots}}$$

3: Results

Simulation Methods

Learning:
In all experiments, each sequence is presented only once.

Recall:
To test recall, we prompt with the first item of the sequence. For recall, performance is measured by how similar the recall and original traces are, averaged across all levels.

Recognition:
To test recognition, we construct a novel sequence, with time-warping and/or per-slice perturbation, from the original and present it to the model. For recognition, in general, there will not be a 1-to-1 correspondence between the items of the original and recalled sequences (due to time-warping). Therefore performance is measured as the similarity between the recognition and original traces but only for the final item of the two sequences.

Experiment 1: Non-Complex Sequence Data Set

- A time-warped and spatially perturbed (one feature per slice changed at random) variant of the original episode above.
- Nearly 100% recognition accuracy (data not shown)

- One of Ten Original Episodes

- Another time-warped variant of the original episode above. No spatial perturbation. Nearly 100% recognition accuracy (data not shown).

Speed-up or Slow-down factors

Input Parameters
- Ten non-complex sequence
- 16 items (time slices) each
- 100 input features
- 10/100 features chosen at random on each slice.
- Two different schedules of non-uniform warping

Total Weights: 934,106

	CMs	Cells/CM	Cells
L5	12	5	60
L4	12	8	96
L3	16	18	288
L2	16	22	352
L1	-	-	100

Recall Acc. (%)	
L5	90.5
L4	92.7
L3	92.8
L2	92.9
L1	92.0

Experiment 2: Complex Sequence Data Set

Comparison between Original and Recalled Trace for Sequence 0

Input Parameters
- Six complex sequence
- 8 items (time slices) each
- 100 input features
- Alphabet of 15 states
- 10% features active on each slice
- 6 unique states
- Non-uniform warping

To visually verify that the traces are the same, compare the original and recalled codes at a given level and time slice (vertical column). An example comparison is shown (blue boxes and connecting line) at time slice 14.

Prompt: When testing recall, we prompt with the first time slice of a sequence.

Total Weights: 240,572

Parameters	CMs	Cells/CM	Cells
L5	9	5	45
L4	9	8	72
L3	12	13	132
L2	16	16	256
L1	-	-	100

Recall Acc. (%)	
L5	100.0
L4	100.0
L3	100.0
L2	100.0
L1	90.3

Experiment 3: Complex Sequence Data Set

Input Parameters
- One complex sequence
- 8 items (time slices)
- 6 unique states
- Non-uniform warping

Original

Recognition

Comparison between the original and recognition traces on the final sequence item, E, which occurs as the 8th time slice in the learning (original) trial, and due to the time-warping, as the 12th time slice during the recognition trial.

- There is a single error at L2 on the first sub-slice when item E presents during recognition.
- It is rectified on the following sub-slice.

Total Weights: 34,917

Parameters	CMs	Cells/CM	Cells
L5	6	5	30
L4	6	5	30
L3	9	9	81
L2	9	9	81
L1	-	-	64

Recall Acc. (%)	
L5	100.0
L4	100.0
L3	100.0
L2	100.0
L1	99.7

4: Summary

Uusitalo et al. (1996) Persistence Data

- Magnetic source imaging (MSI) of evoked responses to brief (70 msec) checkerboard stimulus
- They clustered cortical areas into two groups based on lifetimes
  - Occipital: 0.1–0.6 sec.
  - Temporal, Parietal, & Frontal: 7–30 sec.

- We've provided simulation results of a hierarchical neural network that performs single trial learning of complex sequences and can then recall them when prompted and recognize non-uniformly time-warped and noisy versions of them.
- The two major principles underlying the model are:
  - Its use of a sparse distributed representations. In general, this affords high storage capacity in terms of bits/synapse, although these simulations do not demonstrate that. Secondly, sparse distributed representations, i.e., combinatorial representations, allow an additional representational dimension in that the similarity of two different memory traces can be encoded in the degree of overlap (intersection) between the two traces.
  - Graded persistence such that cells that are late in the 'cortical' hierarchy remain on longer than cells that are early. This allows higher-level codes to become associated with multiple successive underlying codes and leads to robust time-invariant recognition.

Future Work

1. The increasing persistence gradient is crucial to the model's ability to handle time-invariance. However, the model needs modification in order to explain how many sequences that start with the same item or prefix can be mapped into different higher-level codes. We will explore the addition of a working memory mechanism, e.g., the Theta/gamma-based model of Lisman & Idiart (1995) to solve this problem.

References

1. Uusitalo, Samuel J. Williamson, Mika T. Sepp "Dynamical organization of the human visual system revealed by lifetimes of activation traces" (1996) Neuroscience Letters 213 (1996) 149-152.
2. Lisman, J. & Idiart, M.A.P. (1995) "Storage of 7 ± 2 short-term memories in oscillatory subcycles" Science 267:1512-1515.
3. Fuster, J. M. & Alexander, G. E. "Neuron activity related to short-term memory" Science Wash. DC 173: 652-654, 1971.