Lecture 11: Auto-Associative Memory

Instructor: Justin C. Sanchez, PhD
GMS 7795
Section 4837
Delay and State

- Biologic example of maintenance of neuronal activity.
- A single feedforward neuron is not able to maintain outputs (remember membrane reset).
On Recurrency, Dynamics, and Time

- **Feed forward networks:**
  - Information only flows one way
  - One input pattern produces one output
  - No sense of time (or memory of previous state)

- **Recurrency**
  - Nodes connect back to other nodes or themselves
  - Information flow is multidirectional
  - Sense of time and memory of previous state(s)

- An instantaneous connection between the PEs (a product by a weight) leads to infinite looping which is unrealistic and can not be modeled. Delays have therefore to be incorporated either at the PE or at the interconnection level to create time dependencies.
Auto-Associative memory: Can it enhance pattern completion?

See part of a fig. recall the whole fig.

How can cycling activity contribute to short term memory (pattern active until another is presented) or long term memory (persistent synaptic changes)?
Two-stage model of long-term memory formation (Hasselmo).

**Active waking**

- Information from environment
- Rapid encoding
- Intermediate-term episodic representation
  - (Excitatory connections between CA3 pyramids are strengthened that are activated by the event)
  - Reactivation of the stored representations including spread of activity across the strengthened excitatory connections\(^6,^{10}\)
  - Further strengthening of CA3-CA3 and CA3-CA1 connections
  - Slower strengthening of the excitatory associative synaptic connections between neurons representing these features in neocortex

**Quiet waking, SWS**

- Formation of links between these neurons that could be described as semantic or long-term episodic memories in neocortex
Neuropsychological evidence for the two-stage model

- Animal studies: extensive lesions: flat gradient of retrograde amnesia
  partial lesions: temporally graded retrograde amnesia

- Human data: hippocampal damage caused a very long-term retrograde amnesia
  (but: these lesions included subiculum and EC whereas
  very short-term retrograde amnesia occurs when damage is restricted to
  specific hippocampal subfields)

- Thus, consolidation does not necessarily result in formation of links in association
  neocortex, but could instead result in strengthening of representations within
  hippocampus itself or within the EC
Computational evidence for the two-stage model

- Diverse computational models describing the interaction between hippocampus and neocortex when reactivating representations for consolidation

- Key feature of these models:
  → Initial storage of an association in an associative matrix (= component of hippocampus) which requires capacity for rapid synaptic modification in the hippocampus

  → Then, stored representations from this matrix are repetitively retrieved to activate units in association neocortex (this repetitive reactivation results in gradual strengthening of the connections between neurons in the associative neocortex)
Hopfield Networks

- Sub-type of recurrent neural nets
  - Fully recurrent
  - Weights are symmetric
  - Nodes can only be on or off
  - Random updating

- Learning: **Hebb rule** (cells that fire together wire together)
  - Biological equivalent to LTP and LTD

- Can recall a memory, if presented with a corrupt or incomplete version
Discrete Hopfield Model (1982)

- **Architecture:**
  - single layer (units serve as both input and output)
  - nodes are threshold units (bipolar)
  - weights: fully connected, symmetric

\[ w_{ij} = w_{ji} \]

- $x_i$ are external inputs, which may be transient or permanent
Hopfield Nets cont.

- Recurrent networks of non-linear units are generally very hard to analyze. They can behave in many different ways:
  - Settle to a stable state
  - Oscillate
  - Follow chaotic trajectories that cannot be predicted far into the future.

- But Hopfield realized that if the connections are symmetric, there is a global energy function
  - Each “configuration” of the network has an energy.
  - The binary threshold decision rule causes the network to settle to an energy minimum.
**Structure**

- N units (neurons)
- For simplicity of notation, each neuron takes value of only -1 (silent) and 1 (firing)
- Fully connected
- Weight between *ith* neuron and *jth* neuron is given by

\[
W_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^{(\mu)} \xi_j^{(\mu)}
\]

Where \(x^{(m)} = \{ \xi_i^{(\mu)} \mid i = 1, 2, \ldots, N \}\) are patterns to be memorized by the network

Hence the model will remember \(p\) patterns here
Dynamics

After setting the initial state of a Hopfield net, its operation proceeds by repeatedly doing:

pick a node $i$ at random

update it according to the rule

$$s_i(t+1) = \begin{cases} 
1 & \text{if } \sum_j w_{ij} s_j(t) > 0 \\ 
-1 & \text{otherwise} 
\end{cases}$$
The network could update asynchronously or synchronously

**Asynchronously**

Only one unit (neuron) updates its state at a time step.
Randomly pick up a neuron, then update it.

**Synchronously**

All neurons update their states
- Synchronous updates can easily lead to oscillation
Retrieving Memories
We define a quantity called overlap to measure how success is the network

Overlap between $S(t)$ and $x^{(m)}$

$$O(s(t), \xi^{(\mu)}) = 1 - \frac{\sum_{i=1}^{N} |s_i(t) - \xi_i^{(\mu)}|}{N}$$

$s(t) = x^{(m)}$ if and only if $O(s(t), x^{(m)}) = 1$
Hence if a pattern (memory) $x^{(1)}$ is retrieved, we should have

$$O(s(t), \xi^{(\mu)}) = 1$$

Due to symmetry in the Hopfield model, if the model finally stays at either $x^{(m)}$ or $-x^{(m)}$, the stored pattern $x^{(m)}$ is successfully retrieved.
Properties of Hopfield Nets

- **Distributed Representations**
  - Memories are stored across the whole set of units.
  - Different memories correspond to different patterns of activation across the units.

- **Local Asynchronous Control**
  - Each unit makes its own local decision (based on locally available information) about its activation, asynchronously from any other unit.

- **Content-addressable memory** (also known as **associative memory**)
  - A number of patterns can be stored in the network.
  - A pattern can be retrieved simply by specifying part of it, and the network will automatically find the closest stored pattern.

- **Fault Tolerant**
  - The network will continue to function reasonably if a few of the nodes or connections are damaged in some way.
  - What happens is that **all** memories are degraded a bit, rather than losing the whole of some memories.
The operation of a Hopfield net can be interpreted in terms of energy minimization.

By analogy with various physical systems we can define an energy for a Hopfield net by:

\[ E(t) = -\frac{1}{2} \sum_{i,j} w_{ij} s_i(t) s_j(t) \]
Hopfield Nets and Energy Minimization

Any change in state of a node in accordance with the state update rule will cause the energy to decrease or remain unchanged.

It can never increase.
Hopfield Nets and Energy Minimization

The net's energy landscape will then have local minima corresponding to the \( p \) patterns.
Hopfield Nets and Energy Minimization at time zero, initial stimulus

If the network is started off in a state which is a perturbed version of one of the patterns $m$, the network will settle into a state with its activation levels
Hopfield Nets and Energy Minimization at time zero,

\[ \xi(1) \]

= initial stimulus

\[ \xi(2) \]

\[ -\xi(1) \]

\[ -\xi(2) \]
Hopfield Nets and Energy Minimization at time zero,

\[ \xi(1) \]

[diagram]

\[ \xi(2) \]

\[ -\xi(1) \]

\[ -\xi(2) \]

= initial stimulus
Hopfield Nets and Energy Minimization at time zero,

\[ \xi(1) \]

\[ \xi(2) \]

\[ -\xi(1) \]

\[ -\xi(2) \]

= initial stimulus
Example of auto-associative memory

Used to recall a pattern by its noisy or incomplete version.

(pattern completion/pattern recovery)

- A single pattern \( \mathbf{s} = (1, 1, 1, -1) \) is stored (weights computed by Hebbian rule)

\[
\mathbf{W} = \begin{bmatrix}
1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 \\
1 & 1 & 1 & -1 \\
-1 & -1 & -1 & 1
\end{bmatrix}
\]

- Training pat. \( (111-1) \cdot \mathbf{W} = (4\ 4\ 4\ -4) \rightarrow (111-1) \)
- Noisy pat \( (-111-1) \cdot \mathbf{W} = (2\ 2\ 2\ -2) \rightarrow (111-1) \)
- Missing info \( (0\ 0\ 1-1) \cdot \mathbf{W} = (2\ 2\ 2\ -2) \rightarrow (111-1) \)
- More noisy \( (-1-11-1) \cdot \mathbf{W} = (0\ 0\ 0\ 0) \) not recognized
A Hopfield net always makes decisions that reduce the energy.
- This makes it impossible to escape from local minima.

We can use random noise to escape from poor minima.
- Start with a lot of noise so it's easy to cross energy barriers.
- Slowly reduce the noise so that the system ends up in a deep minimum. This is “simulated annealing”.
Are there Hopfield Nets in the Brain??

- The cerebral cortex is full of recurrent connections, and there is solid evidence for Hebbian synapse modification there. Hence, the cerebrum is believed to function as an associative memory.

- Flip-flop figures indicate distributed hopfield-type coding, since we cannot hold both perceptions simultaneously (binding problem)
Assignment

- Read Chapter 8
- Homework on web