Coupled Neural Associative Memories

Amir Hesam Salavati, Amin Karbasi, Amin Shokrollahi

Monday 9 February 15
Puzzle!

Memorize the following images
Memorize the following images
Now answer!
What was the most similar painting to this one?
Now answer!

What was the most similar painting to this one?
What was the most similar painting to this one?
Neural Associative Memory

- Memorize
- Retrieve
Neural Associative Memory

- Memory
- Retrieval
Neural Associative Memory

- Memory
- Recall in presence of noise

Learning
Good noise tolerance
Neural Associative Memory

- **Noise**
- **Robustness**

**Learning**

**Good noise tolerance**

**Large capacity**
Neural Associative Memory

- **Memory**
- **Recall in presence of noise**

**Learning**

**Good noise tolerance**

**Large capacity**

- Artificial neural networks to mimic brain:
  
  [Hopfield, 1982], [McEliece et al., 1987], [Venkatesh et al. 1989], [Komlos et al., 1993], [Lee, 2001], [Muezzinoglu et al. 2003], [Salavati et al. 2011], [Gripon et al., 2011], [Karbasi et al., 2012]
Neural Associative Memory

- Memory
- Retain

Learning
- Good noise tolerance
- Large capacity

• Artificial neural networks to mimic brain:
  [Hopfield, 1982], [McEliece et al., 1987], [Venkatesh et al. 1989],
  [Komlos et al., 1993], [Lee, 2001], [Muezzinoglu et al. 2003],
  [Salavati et al. 2011], [Gripon et al., 2011], [Karbasi et al., 2012]
Traditional Approach
Traditional Approach

- Design a network to memorize any set of random patterns
Traditional Approach

- Design a network to memorize any set of random patterns

[Hopfield, 1982], [McEliece et al., 1987], [Venkatesh et al. 1989], [Komlos et al., 1993], [Lee, 2001], [Muezzinoglu et al. 2003]

Problem: versatility causes low capacity

Out of $2^n$ possible binary vectors of length $n$, only $O(n)$ can be memorized
Puzzle, Again!

Now memorize these images:
Puzzle, Again!

Now memorize these images:
What was the most similar painting to this one?
Structured Patterns

- Structure patterns increase the capacity
  - $O(n^2)$ [Gripon & Berrou 2011]
  - $O(a^n)$ with $a > 1$ [Kumar et al. 2011]
Structured Patterns

- Structure provides a way to increase capacity
  - $O(n^2)$ [Gripon & Berrou 2011]
  - $O(a^n)$ with $a > 1$ [Kumar et al. 2011]

Learning: ✓
Good noise tolerance: ×
Large capacity: ✓
In This Talk...

- Improve noise tolerance
- Some history
- New perspective from convolution to coupled

- Simulation results
- ...
The Model & Some History
Neural Model

• Patterns
  • Vectors of length
  • Integer values and non-negative (firing rate)
    • e.g. quantized grey level values
Neural Model

- Patterns
  - Values of length
  - Integer values and non-negative (firing rate)
    - e.g. quantized grey level values

- Strong local correlations: each sub-pattern comes from a subspace

vs.
The Learning Process

- Learn the dual vector(s) orthogonal to the sub-patterns
- Look for sparse or orthogonal vectors
The Learning Process

- Learn the dual vectors orthogonal to the sub-patterns
- Look for sparse orthogonal vectors
The Learning Process

- Learn the dual vector(s) orthogonal to the sub-patterns
- Look for sparse or orthogonal vectors

\[ y_1 \]

\[ x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_n \]
The Learning Process

- Learn the dual vector(s) orthogonal to the sub-patterns.
- Look for sparse orthogonal vectors.

\[
\begin{align*}
&x_1, x_2, x_3, x_4, \ldots, x_n \\
&y_1, y_2
\end{align*}
\]
The Learning Process

- Learn the dual vector(s) orthogonal to the sub-patterns
- Look for sparse or orthogonal vectors
The Learning Process

- Learn the dual vectors orthogonal to the sub-patterns
- Look for sparse or orthogonal vectors

All in all, we have a parity-check graph!
The Recall Phase

- **Theorem 1:** Each block could correct 1 error reliably.

[1] *Iterative learning and denoising in convolutional neural associative memories*

A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem 1**: Each block could correct 1 error reliably.
- However, the overlap among blocks helps achieve correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem [1]**: Each block could correct 1 error reliably.

- However, the overlap among blocks helps achieve correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Thm. 1:** Each block could correct one error reliably.

- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*  
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem 1**: Each block could correct 1 error reliably.

- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem [1]**: Each block could correct 1 error reliably.
- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*  
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem**: Each block could correct one error reliably.
- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] Iterative learning and denoising in convolutional neural associative memories
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem**: Each block could correct 1 error reliably.
- However, the overlap among blocks help achieve correct correction of a linear fraction of errors [1].

[1] *Iterative learning and denoising in convolutional neural associative memories*

A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- **Theorem**: Each block could correct 1 error reliably.

- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] Iterative learning and denoising in convolutional neural associative memories

A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
The Recall Phase

- Theorem 1: Each block could correct 1 error reliably.
- However, the overlap among blocks help achieve correction of a linear fraction of errors [1].

[1] Iterative learning and denoising in convolutional neural associative memories
A. Karbasi, A. H. Salavati, A. Shokrollahi, ICML 2013
Relations to Peeling Decoder
Relations to Peeling Decoder
Relations to Peeling Decoder

- Very similar to the Peeling Decoder over the following graph.
Coupled Associative Memories
Coupling Neural Graphs

- Same decoding principle over clusters and planes
Coupling Neural Graphs

- Same decoding principle: sequential over clusters and then planes

- Side information: freeze border neurons to the correct value
Coupling Neural Graphs

- Same decoding principle: sequential over clusters and then planes

- Side information: freeze border neurons to the correct value

- Parameters: Monday 9 February 15
Coupling Neural Graphs

- Same decoding principle: sequential over clusters and then planes.

- Satisfaction: freezing border neurons to the correct value.

- Parameters:
  - $D$: number of planes
Coupling Neural Graphs

- Same decoding principle: sequential over clusters and then planes

- Side information: freezing boundary neurons to the correct value

- Parameters:
  - $D$: number of planes
  - $L$: number of clusters in each plane
Coupling Neural Graphs

- Same decoding principle: sequences over clusters and then planes

- Side information: freezing border neurons to the correct value

- Parameters:
  - \( D \): number of planes
  - \( L \): number of clusters in each plane
  - \( \Omega \): coupling window
Biological Appeals

- Satifandsion from the cognitive:

  "the..." (choke)
Biological Appeals

- Satirband from the cognitive level:
  "the ants" (cat or bat?)

- Similar "spatial connections" in mammals

  Mada et al., Cognitive computing, Communications of the ACM 2011.
• Technical borrow from [2]

- Technical tools borrowed from [2]

Date

- $p_e$: "channel" error probability
- $z(t)$: average probability of error at time $t$
- $p_e^*$: maximum $p_e$ for which the uncoupled system is successful

• Technical approach [2]

Date

• $p_e$: "channel" error probability

• $z(t)$: average probability of error in iteration

• $p_e^\dagger$: maximum $p_e$ for which the uncoupled system is successful

We characterized $U(z; p_e)$ through the inequality

$$U'(z; p_e) > 0 \text{ for } p_e < p_e^\dagger$$

Performance Analysis

- Technical tools borrowed from [2]

Data

- $p_e$: “channel” error probability
- $z(t)$: average probability of error induced
- $p_e^*$: maximum $p_e$ for which the uncoupled system is successful

We have defined $U(z; p_e)$ to have the property

$$U'(z; p_e) > 0 \text{ for } p_e < p_e^*$$

Define $p_e^* < p_e^*$ to be the maximum $p_e$ for which

$$\min_z U(z; p_e) > 0$$

Results
Error Correction Performance

- Theorem: if the coupling window is large enough, then the coupled system converges to the correct memorized pattern for all error probabilities $p_e < p_e^*$. 
Error Correction Performance

- Theorem: if the coupling window is large enough, then the coupled system converges to the correct memorized pattern for all error probabilities $p_e < p_e^*$. 

  - Note that since $p_e^* < p_e^*$, this means that the coupled system outperforms the uncoupled system.
Error Correction Performance

- Theorem: if the coupling window is large enough, then the coupled system converges to the correct memorized pattern for all error probabilities \( p_e < p_e^* \).

- Note that since \( p_e^\dagger < p_e^* \) this means that the coupled system outperforms the uncoupled system.

- The lower bound for \( \Omega \) provides a sufficient condition.
Simulations

- Pattern error probability vs. initial error rate

Figure 3: The final pattern error probability for the constrained and unconstrained coupled neural systems.

Once finished, we declare failure if the output of the algorithm, $\hat{x}$, is not equal to the pattern $x$ (assumed to be the all-zero vector).

Figure 4 illustrates how the potential function for uncoupled systems behaves as a function of $z$ and for various values of $p_e$.

Note that for $p_e \approx p^\ast e$, the minimum value of potential reaches zero, i.e. $E(p^\ast e) = 0$, and for $p_e > p^\ast e$ the potential becomes negative for large values of $z$.

TABLE I: The thresholds for the uncoupled ($p^\ast e$) and coupled ($p^\ast e$) systems.

<table>
<thead>
<tr>
<th>$e$</th>
<th>$p^\ast e$</th>
<th>$p^\ast e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.078</td>
<td>0.114</td>
</tr>
<tr>
<td>2</td>
<td>0.197</td>
<td>0.394</td>
</tr>
</tbody>
</table>

IX. CONCLUSIONS

In this paper, we proposed a novel architecture for neural associative memories. The proposed model comprises a set of neural planes with sparsely connected overlapping clusters. Furthermore, planes are sparsely connected together as well.

Given the similarity of the suggested framework to spatially-coupled codes, we employed recent developments in analyzing these codes to investigate the performance of our proposed neural algorithm. We also presented numerical simulations that lend additional support to the theoretical analysis. We derived two thresholds on the maximum initial bit error probability that can be corrected by the proposed algorithm with probability close to 1. Using simulations, we confirmed that there is a good match between the thresholds derived theoretically and those obtained in practice.

Given that our main interest in this paper was the performance of the error correcting algorithm in the recall phase, we did not address the learning phase here. However, we are currently in the middle of applying the learning method in [6] to a database of natural images to assess the performance of the recall algorithm in this real-world setup as well.
Simulations

- **Results indices**

- **Theoretical results**

<table>
<thead>
<tr>
<th>$e$</th>
<th>$p_e^*$</th>
<th>$p_e^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.078</td>
<td>0.114</td>
</tr>
<tr>
<td>2</td>
<td>0.197</td>
<td>0.394</td>
</tr>
</tbody>
</table>
Ongoing Work
Internal Noise Helps!
Internal Noise Helps!

- The neurons in our model were perfect, i.e., deterministic
Internal Noise Helps!

- The neurons were perfect, i.e., deterministic.
- But real neurons are susceptible to internal noise.
- So what happens if we introduce internal noise?
Internal Noise Helps!

- The neurons in our model were perfect, i.e., deterministic.
- But real neurons are susceptible to internal noise.
- So what happens if we introduce internal noise?

Rather counterintuitively, internal noise improves the error correction performance!

The network achieves better thresholds in presence of internal noise.

Noise-Enhanced Associative Memories
A. Karbasi, A. H. Salavati, A. Shokrollahi, L. R. Varshney To appear in NIPS 2013

ITW 2013, Seville, Spain

Monday 9 February 15
Backup Slides
There exists a data set $X$ with vectors of length $n$ such that $C = a^k$, with $a \geq 2$, and $k = \text{rank}(X) = O(n)$. 