Progress Report
2 to 13 September 2010

Amir Hesam Salavati
E-mail: hesam.salavati@epfl.ch

Supervisor: Prof. Amin Shokrollahi
E-mail: amin.shokrollahi@epfl.ch

Algorithmics Laboratory (ALGO)
Ecole Polytechnique Federale de Lausanne (EPFL)

September 13, 2010

1 Introduction

In this report, I am going to analyze the coded Hopfield network as proposed by Berrou in [1]. The structure of coded Hopfield networks is discussed in some details in my last report. I start with the local decoding procedure and calculate its probability of error. Then, I will consider the global probability of error doing one iteration of decoding.

2 Local Decoding

The local decoder is in charge of determining which neuron (called fanal in here) should be active given the input pattern. Hence, an error occurs if there are more than one fanal active or the wrong fanal is selected. A fanal is selected by the decoder if its weighted input sum is maximum compared to other fanals. Denote this weighted sum by $z_1, \ldots, z_J$, where J is the number of fanals. Now assume the all one message was sent, i.e. $y_1 = \ldots = y_\kappa = 1$, then the correct fanal is the first one. In this case, the probability of correct local decoding, P_c^L is:

$$P_c^L = \prod_{j=2}^{J} Pr\{z_1 > z_j\} \tag{1}$$

Here we have assumed messages, and hence their corresponding fanals, are independent of each other. Let $P_j = Pr\{z_1 > z_j\}$. Then:

$$P_j = Pr\{z_1 > z_j\} = Pr\{\sum_{i=1}^{\kappa} g_{i1} y_i > \sum_{i=1}^{\kappa} g_{ij} y_i\} = Pr\{\sum_{i=1}^{\kappa} (g_{i1} - g_{ij}) y_i > 0\} \tag{2}$$

Recall that g_{ij} is the weight between the input neuron i, y_i, and the fanal j. Furthermore, these weights are determined according the message bits. In other words, if fanal j corresponds to message m_j, then g_{ij} is the i^{th} bit of m_j. Since we have assumed that the first fanal corresponds to the all one message, then $g_{i1} = 1$ for $i = \{1, \ldots, \kappa\}$.

At this point, we assume g_{ij} is chosen uniformly at random. **The validity of this assumption needs to be checked, specially because of the facts that it involves the way messages are selected.** Nevertheless, if we assume the uniform distribution assumption, then

$$g_{i1} - g_{ij} = 1 - g_{ij} = \begin{cases} 0, & \text{with probability } 1/2 \\ 2, & \text{with probability } 1/2 \end{cases}$$

As a result, we obtain:

$$(g_{i1} - g_{ij})y_i = \begin{cases} 0, & \text{w.p. } 1/2 \\ 2, & \text{w.p. } (1-\epsilon)/2 \\ -2, & \text{w.p. } \epsilon/2 \end{cases} \tag{3}$$

where ϵ is the probability of error in the BSC channel.

Therefore, the sum $\sum_{i=1}^{\kappa}(g_{i1}-g_{ij})y_i$ is the sum of κ i.i.d. random variables with mean $\mu = 1-2\epsilon$ and variance $\sigma^2 = 2-\mu^2$. Thus, equation (2) simplifies to:

$$\begin{aligned} P_j = Pr\{\sum_{i=1}^{\kappa}(g_{i1} - g_{ij})y_i > 0\} &= Q\left(\frac{-\mu\kappa}{\sqrt{\kappa(2-\mu^2)}}\right) \\ &= Q\left(\sqrt{\frac{\kappa(2\epsilon-1)^2}{2-(2\epsilon-1)^2}}\right) \end{aligned} \tag{4}$$

In which Q is the standard Q-function. By replacing equation (4) int (1) we obtain the probability of local decoding error as follows:

$$P_e^L = 1 - \prod_{j=2}^{J} Pr\{z_1 > z_j\} = 1 - \left[Q\left(\sqrt{\frac{\kappa(2\epsilon-1)^2}{2-(2\epsilon-1)^2}}\right)\right]^{J-1} \tag{5}$$

3 Global Decoding

In the global decoding process, the decoder tries to determine the most probable pattern based on the result of the local decoders inside neural clusters. More specifically, the global decoding step is as follows:

$$v^{\ell}(n_{ij}) = \sum_{i'=1}^{C}\sum_{j'=1}^{J} w_{(ij)(i'j')}v^{\ell}(n_{i'j'}) + \gamma v^{\ell-1}(n_{ij}) \tag{6}$$

In the above equation, n_{ij} is the j^{th} fanal of the i^{th} cluster, $v^\ell(n_{ij})$ is the output of n_{ij} at round ℓ, C is the number of clusters, $w_{(ij)(i'j')}$ is the weight (binary) between n_{ij} and $n_{i'j'}$ and finally γ is a constant (greater than or equal to 1) which simulates the memory effect of neurons. The weights $w_{(ij)(i'j')}$ are determined according the message patterns as explained in the previous reports.

In each round, the output of all neurons are computed based on equation (6) and then the fanal in each cluster that has the *maximum* output value is set to fire, i.e. its output is made equal to 1 while other neurons inside the same cluster remain silent. Hence, an error occurs if inside a cluster, a wrong fanal or more than one fanals fire.

For simplicity, suppose we have four clusters and the transmitted pattern is the one which involves the first fanal of each cluster, i.e. n_{11}, n_{21}, n_{31} and n_{41}. In other words, we have $w_{(ij)(i'j')} = 1$ for these neurons. To see how the global decoder handles errors, let's assume after the local decoding step, only one cluster is erroneous. More specifically, suppose in the first cluster, instead of n_{11}, we had n_{12} to fire. Therefore, in the next round and based on equation (6), we have:

$$\begin{aligned}
v^\ell(n_{11}) &= 3\\
v^\ell(n_{12}) &= w_{(12)(21)} + w_{(12)(31)} + w_{(12)(41)} + \gamma\\
v^\ell(n_{21}) &= 2 + \gamma + w_{(12)(21)}\\
v^\ell(n_{31}) &= 2 + \gamma + w_{(12)(31)}\\
v^\ell(n_{41}) &= 2 + \gamma + w_{(12)(41)}
\end{aligned}$$

If we denote the probability of global decoding error in round ℓ by $P_e^G(\ell)$, then for *a given cluster* have:

$$\begin{aligned}
P_e^G(\ell) = &\, Pr\{\text{Correct in round } \ell-1 \text{ and wrong in this round}\} \times P_c^L(\ell-1)\\
&+ Pr\{\text{Wrong in round } \ell-1 \text{ and wrong in this round}\} \times P_e^L(\ell-1) \quad (7)
\end{aligned}$$

If we denote $Pr\{\text{Correct in round } \ell-1 \text{ and wrong in this round}\}$ by P_1 and $Pr\{\text{Wrong in round } \ell-1 \text{ and wrong in this round}\}$ by P_2, then P_1 corresponds to the case that n_{i1} fires at round $\ell-1$ but mistakenly made silent in round ℓ. Likewise, P_2 corresponds to the case that n_{i1} was mistakenly made silent both in rounds $\ell-1$ and ℓ. As a result and according to the values of

$v(n_{ij})$ displayed above we will have:

$$
\begin{aligned}
1 - P_1 = & Pr\{2 + \gamma + w_{(12)(21)} > w_{(12)(2j)} + w_{(31)(2j)} + w_{(41)(2j)}, \forall\ j \neq 1\} \\
& \times Pr\{2 + \gamma + w_{(12)(31)} > w_{(12)(3j)} + w_{(21)(3j)} + w_{(41)(3j)}, \forall\ j \neq 1\} \\
& \times Pr\{2 + \gamma + w_{(12)(41)} > w_{(12)(4j)} + w_{(21)(4j)} + w_{(31)(4j)}, \forall\ j \neq 1\} \quad (8)
\end{aligned}
$$

$$
P_2 = Pr\{3 \leqslant w_{(12)(21)} + w_{(12)(31)} + w_{(12)(41)} + \gamma\} \quad (9)
$$

To analyze equations (8) and (9), let's assume there is a link between any two fanals of different clusters in the graph with probability q (this is what called graph density in [1]). Then, since $\gamma \geqslant 1$ we can write:

$$
\begin{aligned}
& Pr\{2 + \gamma + w_{(12)(21)} > w_{(12)(2j)} + w_{(31)(2j)} + w_{(41)(2j)}, \forall\ j \neq 1\} \\
= & Pr\{w_{(12)(21)} = 1\} \times 1 + Pr\{w_{(12)(21)} \neq 1\} \times Pr\{2 + \gamma > w_{(12)(2j)} + w_{(31)(2j)} + w_{(41)(2j)}, \forall\ j \neq 1\} \\
= & q + (1 - q) \prod_{j=2}^{J} Pr\{2 + \gamma > w_{(12)(2j)} + w_{(31)(2j)} + w_{(41)(2j)}\}
\end{aligned}
$$

If we assume $\gamma = 1$ as in the original paper, then we will have:

$$
Pr\{2 + \gamma + w_{(12)(21)} > w_{(12)(2j)} + w_{(31)(2j)} + w_{(41)(2j)}, \forall\ j \neq 1\} = q + (1 - q) \left(1 - q^3\right)^{J-1}
$$

Following a similar approach for all the clusters, we get:

$$
P_1 = P_c^L(\ell - 1) \left[1 - (q + (1 - q) \left(1 - q^3\right)^{J-1})\right] \quad (10)
$$

In a similar fashion and by assuming $\gamma = 1$, we can write P_2 as:

$$
\begin{aligned}
P_2 &= P_e^L(\ell - 1) Pr\{2 \leqslant w_{(12)(21)} + w_{(12)(31)} + w_{(12)(41)}\} \\
&= P_e^L(\ell - 1) \left[q^3 + \binom{3}{2} (1 - q) q^2\right] \quad (11)
\end{aligned}
$$

Replacing equations (10) and (11) into (7), we obtain the global decoding error probability as follows:

$$
\begin{aligned}
P_e^G(\ell) = & P_c^L(\ell - 1) \left[1 - (q + (1 - q) \left(1 - q^3\right)^{J-1})\right] \\
& + P_e^L(\ell - 1) \left[q^3 + \binom{3}{2} (1 - q) q^2\right] \quad (12)
\end{aligned}
$$

where $P_e^L(\ell - 1) = 1 - P_c^L(\ell - 1)$ is the local decoding error probability in round $(\ell - 1)$ and could be obtained from equation (5). **However, note that extending the result of equation (5) to rounds $\ell > 1$ requires extra care as this equation was derived for the first iteration.**

4 Future Works

So far, I have tried to analyze the behavior of the Berrou's coded Hopfield network in a simple scenario, i.e. when we have only 4 clusters out of which one of them returns error. I have already checked the global decoding probability of error with MATLAB and it drops quite rapidly. Extending this approach to the general case would be among my next steps. I will also try to simulate the performance of the proposed network in terms of probability of error vs. the size of the graph.

References

[1] V. Gripon, C. Berrou, Sparse Neural Networks with Large Learning Diversity, Submitted to IEEE Transaction on Neural Networks.