Extraction of Coherent Information from Non-Overlapping Receptive Fields

Dario Floreano

AREA Science Park, Cognitive Technology Laboratory, Padriciano 99, I-34012 Trieste, Italy

Abstract. It has been suggested that long-range lateral connections in the cortex play a contextual role in that they modulate the gain of the response to primary receptive field input. In the first part of this paper I show that a network with a set of such pre-wired connections has a short-term dynamics that enhances and stabilizes coherent information defined across multiple, non-overlapping receptive fields. In the second part, I suggest a simple Hebbian rule that can develop the required pattern of synaptic strengths and describe two simulations where the networks discover information that is defined only by its coherence across receptive fields.

1 A contextual role for long-range lateral connections

Neurophysiological and pharmacological experiments have indicated that long-range lateral connections in the cortex have a modulatory effect on the postsynaptic cell [3]. Their action could be described by a mechanism that increases or decreases the gain of the cell response to the receptive field input, but cannot alter the feature that is transmitted by the receptive field of the cell. Since these cortico-cortical connections extend from 200μm up to 6mm, they connect groups of neurons with non-overlapping receptive fields. Each group is internally organized in a fairly similar fashion as a local circuit of neurons with largely-overlapping receptive fields and inhibitory interneurons, such as the orientation-selective hypercolumns in the visual cortex. In this paper such a local circuit will be called a *processor*. Modulatory long-range connections between processors could serve various purposes. For example, they could synchronize activity of processors responding to similar features distributed across the input surface and make them stand out from background activity. Since they affect postsynaptic activity, they could also play a role in the development of the primary receptive fields of the individual processors that they connect.

A continuous activation function has been proposed [5] that combines driving signals from the primary receptive fields with modulatory signals from lateral processing units . The output of the processor is a bipolar value {-1, +1} whose probability is given by filtering the activation strength through a sigmoid function. The activation function intends to capture the main biological properties of the interaction between the two types of signals: *a)* the sign of the output is determined solely by the primary receptive fields, *b)* the strength of the activation is increased when the sign of the integrated modulatory signal agrees with

the sign of the integrated driving signal, and *c)* it is reduced when the sign of the integrated modulatory signal is in contrast with the sign of the integrated driving signal; *d)* the activation function is equivalent to a sigmoid function when the modulatory signal is nil.

$$\mathcal{A}(d_i, m_i) = (1/2)d_i(1 + \exp(2d_i m_i)) \tag{1}$$

where $d_i = \sum_{j=1}^{n} x_j w_{ij}^d$ is the integrated driving signal from the receptive field components x_j and $m_i = \sum_{k=1,k\neq i}^{n} y_k w_{ik}^m$ is the integrated modulatory signal from the other processors y_k. The output probability $P(y_i = 1)$ is then computed by passing $\mathcal{A}$ through a sigmoid function, or through a tanh function for a mean-field approximation.

The modulatory connections provide *contextual guidance* in that they modulate the postsynaptic activity according to what is computed by the other processors. It has also been suggested that contextual – or modulatory – connections might be used to extract information that is coherent across the receptive fields of several processors. The goal of each processor is then the maximization of the three-way mutual information between its own output, the receptive field input, and the contextual field input. Learning rules can be derived by performing gradient ascent on this objective function [5]. Similar goals have been explored also with other learning schemes and architectures [1, 7].

Using this approach, section 2 shows that a network of such processors with pre-wired contextual connections can dynamically enhance and stabilize a noisy signal. Section 3 suggests a simple Hebbian rule that can provide the required functionality.

2 Enhancement and stabilization of noisy signals

Consider a set of 100 processors organized as a square matrix (Figure 1, left).

Each processor receives a single primary receptive field input from a corresponding element in the input surface where a subset of 4 x 4 elements have a value of 0.6 with added random noise from a uniform distribution in the range ± 0.3 and all the remaining elements have a random value in the ± 0.6 range. Only the processors that receive a primary signal from the central 4 x 4 input elements are mutually linked by contextual connections which have a synaptic value of 1. The output of each processor is updated for several iterations: at each iteration the values of the input elements are calculated anew. Note that intial output activity, before contextual modulation, is less than input (Figure 1, step 1). After a single iteration the processors within the 4 x 4 inner matrix become fully active (Figure 1, step 2) and maintain the activity level despite strong fluctuations at the input level by sustaining each other through the contextual signals (Figure 1, step 3). Similar behaviours are observed when the input surface consists of ambiguous figures analogous to the Necker cube or to the Rubin vase and inhibitory contextual connections are added between processors connected

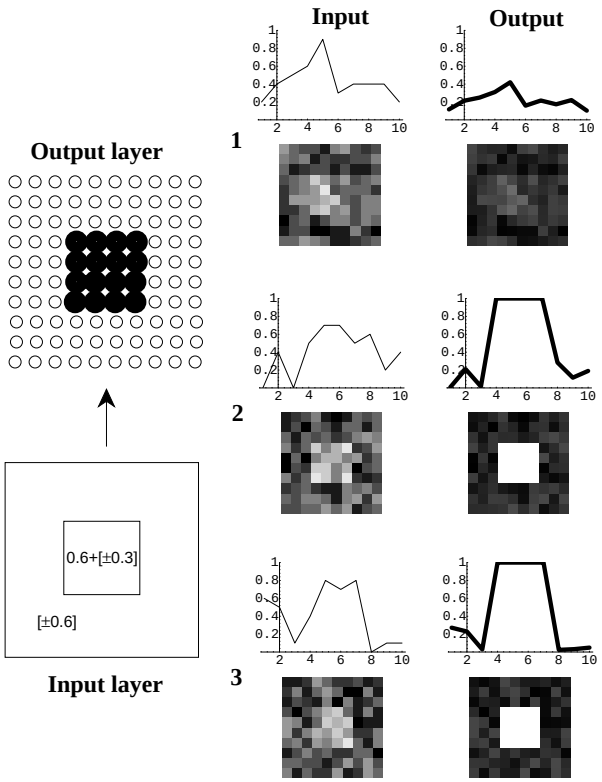

Fig. 1. Left: Architecture of the network. Each processor receives a single primary receptive field input from a corresponding element in the input surface. Processors drawn in bold face are mutually linked by excitatory contextual connections. **Right**: Input and output activity (the absolute value of $\tanh[\mathcal{A}(d_i, m_i)]$) for three iterations; at each iteration the input elements and the processor outputs are computed anew. Step 1 shows intial output activity before the contribution of the contextual signal. For each iteration are shown the complete matrices of input and output activity levels (white and black meaning respectively fully active and completely inactive) and graphs correponding to a slice through the middle row.

to different input subgroups: in this case the network can shift between alternative interpretations of the image and maintain them for a number of iterations [2, 6].

The dynamic properties of this model are potentially useful in multi-stage levels of processing. Appropriate configurations of lateral connections can segregate and enhance a noisy pattern from the background according to its spatial and/or temporal coherence, thus facilitate processing at successive stages. This functionality is also similar to the grouping properties of synchronizing connec-

tions [6].

3 Adaptation of receptive and contextual fields

This section introduces a local Hebbian rule that develops the pattern of synaptic configuration assumed in the model above. Since learning is done on-line, the output of each processor is computed using the mean field approximation

$$y_i = \tanh\left[\mathcal{A}(d_i, m_i)\right] \tag{2}$$

Both the receptive field weights and the contextual field weights are simultaneously adapted by the same learning rule. The main idea is that synaptic weights should be adapted when postsynaptic activity is boosted by the contextual signal, and gradually decay when postsynaptic activity is dampened by the contextual signal. A single binary variable M_i describes the contextual effect on processor i after L iterations of activity ($L = 3$ in the experiments described below)

$$M_i = \begin{cases} 1 \text{ if } |y_i^t| - |y_i^{t-L}| \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Synaptic modification occurs on a slower time scale: all the weights in the network (receptive field and contextual field weights) are simultaneously updated after computation of short-term dynamics according to the same learning rule

$$w_{i,j}^t = w_{i,j}^{t-1} + \eta \underbrace{y_i \left(x_j \overline{y_i x_j} - y_i w_{i,j}^{t-1}\right) M_i}_{\text{adaptation}} \underbrace{- (1 - M_i) \overline{y_i x_j} w_{i,j}^{t-1}}_{\text{decay}} \tag{4}$$

where η is the learning rate, x_j is the preynaptic activity, and $\overline{y_i x_j}$ is a moving average of pre - and postsynaptic activity over a restricted time window (5 learning cycles in the experiments described below). The adaptation component of learning is similar to Oja's learning rule for extraction of the first principal component [4] and makes sure that the synaptic weight vectors tend to length 1. Both the adaptation and the decay phase depend on the average correlation between pre- and postsynaptic activity $\overline{y_i x_j}$: the stronger the correlation, the faster the weights are adapted if boosting of postsynaptic activity has occurred ($M_i = 1$) and the faster they decay to zero if dampening of postsynaptic activity has occurred ($M_i = 0$).

Consider a network with two processors that have non-overlapping receptive fields and are mutually linked by contextual connections. The primary input to each processor is a random vector of three elements that can take bipolar values {-1, +1} with equal probability; however, the sign of the first input element is correlated with the sign of the same input element in the other processor. Both receptive field weights w_{ij}^d and contextual field weights w_{ij}^m are initialized to random values in the ± 0.001 range, and the learning rate η is set to 0.1. Both the processors learn to signal the sign of the first element in their own input. This is reflected by the pattern of connection strengths (Figure 2, **a**). Only the synaptic weights corresponding to the input element correlated across processors are

strengthened. The contextual connections between the two processors are also strengthened at the same time. The small fluctuations of the remaining connections reflect temporary weak correlations between the other input elements and, although they are too small to affect the response of the processors, they can be easily reduced by widening the time window over which $\overline{y_i x_j}$ is computed. If the processors do not share correlated information in their input or such information ceases to exist over time, the contextual connections would gradually decay to zero and each processor would transmit the most informative component of its own receptive field. The same algorithm has also been used to extract higher order information defined only across processors. Each receptive field input can be visualized as a 2 x 2 square matrix whose entries r_{ij} take bipolar values {-1, +1} with equal probability (Figure 2, **b**, left side). Therefore, within each of the two streams of processing all possible inputs occur with equal probability. The higher-order input variable correlated across the two streams is the sign of the "horizontal edge" $E_H = \left(\sum_j r_{1j} - \sum_j r_{2j}\right)$ which is the sign of the difference between the sums of the two row components [5]. Both processors learn to signal the sign of the horizontal edge in less than 450 training cycles (all the training parameters are the same employed in the previous experiments). The final receptive field weights reflect the structure of the variable correlated across processors (Figure 2, **b**, right side).

Summarising, the model outlined in this paper is capable of extracting coeherent information defined across non-overlapping receptive fields, and of developing the appropriate pattern of contextual connections. Since the learning rule includes a normalizing factor, the postsynaptic activation is never saturated and the modulatory signals can generate the short-term dynamics described in section 2.

References

1. S. Becker and G. E. Hinton. Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, vol. 355:pp. 161–163, 1992.
2. D. Floreano, J. Kay, and W. A. Phillips. A computational theory of learning visual features via contextual guidance. European Conference on Visual Perception, August, Tübingen, 1995.
3. J. A. Hirsch and C. D. Gilbert. Synaptic physiology of horizontal connections in cat's visual cortex. *Journal of Neuroscience*, vol. 1:pp. 1800–1809, 1991.
4. E. Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, vol. 15:pp. 267–273, 1982.
5. W. A. Phillips, J. Kay, and D. M. Smyth. The discovery of structure by multi-stream networks of local processors with contextual guidance. *Network*, vol. 6:pp. 225–246, 1995.
6. W. A. Phillips and W. Singer. In search of common foundations for cortical computation. Submitted to *Brain and Behavioural Science*, 1995.
7. J. Schmidhuber and D. Prelinger. Discovering predictable classifications. *Neural Computation*, vol. 4:pp. 625–635, 1993.

This article was processed using the LaTeX macro package with LLNCS style

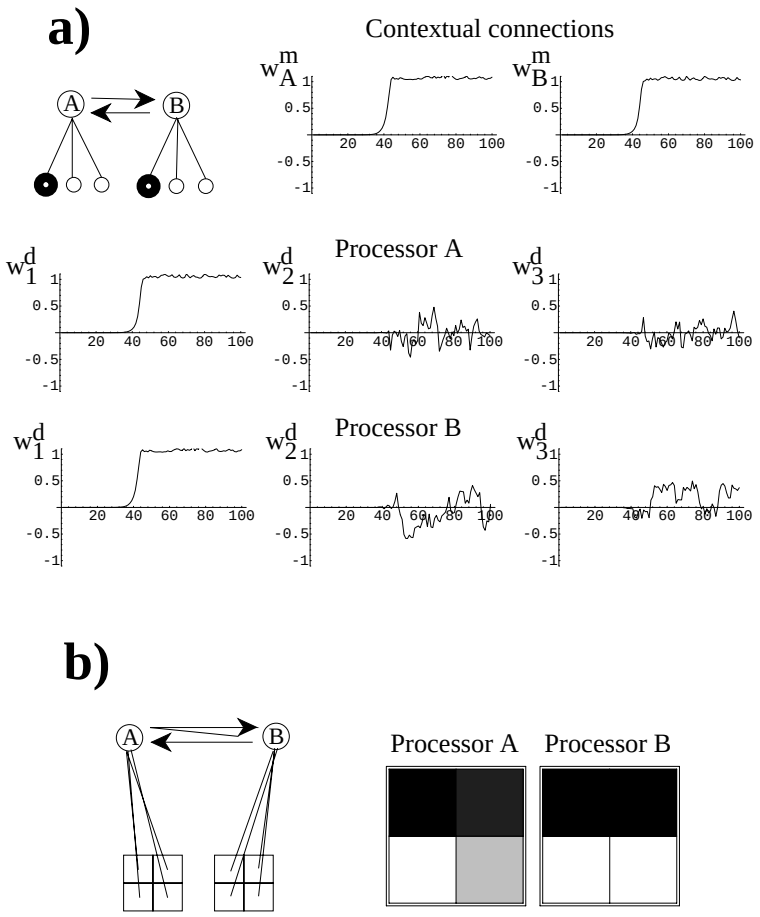

Fig. 2. Extraction of coherent information when there is no structure within the receptive field input of the processors. **a**: Each processor receives primary input from three random bipolar units, but the sign of the first unit (evidentiated in bold face) is correlated across receptive fields. Development of synaptic strengths during learning (data are plotted every 10 cycles): both contextual weights and receptive field weights learn at the same time to signal the sign of the correlated variable. **b**: Architecture and final receptive field weights of a network that has discovered the sign of the horizontal edge variable defined only across processors. Black is inhibitory and white is excitatory (the gray level stands for a small excitatory weight).