An Internal Teacher

Dario Floreano
MANTRA Center for Neural Computation
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
floreano@di.epfl.ch
http://diwww.epfl.ch/lami/team/floreano

Abstract

Contextual signals might supervise discovery of coherently varying information between cortical modules computing different functions of their receptive field input. This hypothesis is explored in two sets of computational experiments, one studying the effects on learning of long-range unidirectional contextual signals mediated by intervening processors, and the other showing contextually supervised discovery of a high-order variable in a multi-layer network.

1 Supervision and biological plausibility

Supervised models represent a successful paradigm of learning for artificial neural networks. Not only such models have been applied to a wide range of engineering problems, but they have also been used to infer computational properties of living brains (e.g., [4]). However, the requirement of desired values for output neurons (provided by an "external teacher"), often coupled with back-propagation of error signals along the same synaptic connections [3], raises serious doubts about the biological plausibility of neural models of supervised learning [2].

The theoretical framework and the computational model described by Phillips and Singer might now provide a conciliatory solution which combines inter-module supervision with a biologically plausible learning scheme. Contextual signals between different processors computing different functions of their inputs might play the role of an *internal teacher*. The idea is that spatial and/or temporal correlation of events could be exploited by the neural system to guide acquisition of a novel processing ability. In order to test this hypothesis, there are at least three constraints which must be satisfied: i) if supervision takes place between processors computing different functions (which might thus be located far apart in the cortical tissue), it is necessary to check whether contextual guidance can have effects at long distance (longer than that implied by the average extension of lateral intra-cortical connections), for example through the mediation of other intervening processors; ii) since supervision makes sense only when at least one of the interconnected cortical processors cannot discover the appropriate function, the *internal teacher* hypothesis should be tested on a computationally "difficult" function, such as a non-linear transformation of the input; iii) although one could imagine a situation where an already developed neural module supervises learning of another module, the model should also account for the situation when both modules learn together.

2 Long-range effects of contextual signals

The hypothesis that contextual signals might function as an *internal teacher* was tested in a set of computational experiments guided by the above constraints and based on the mathematical model outlined in the target article. An important feature of these experiments is that contextual connections between different processors (or modules, if they are composed of more than one processor) were unidirectional, whereas in all the other experiments described in the target article contextual connections were reciprocal and symmetrical. The first experiment was aimed at testing the first constraint, namely whether a contextual signal traveling through several processors with non-overlapping receptive fields could guide discovery of the same feature in all processors. The architecture was composed of three processors that were presented with a horizontal edge contrast whose sign was correlated across processors (see top of figure). The structure of the input for each processor can be easily understood if we visualize the receptive field of each module as a 2x3 matrix whose entries can take random bipolar values -1 or +1. The sign of the contrast edge was given by the difference between the two sums of the row components. For every pattern presentation, each processor received a new random input, but the sign of the edge contrast was correlated across processors. Although this arrangement might look a little bit artificial, it guaranteed that a simple Infomax approach could not discover the information correlated across processors. The three processors were arranged on a line, each one receiving a contextual signal through a unidirectional connection from the preceding processor; the contextual signal to the first processor, which in this experiment was supposed to come from an already developed neural module, was a bit whose value alternated together with the sign of the edge contrast presented as input to the network. Starting from initially-random weight strengths for both RF and CF connections, in less than 500 learning cycles all processors learned to signal the sign of the edge contrast even though the contextual signal to the last two processors were mediated by intervening processors (the graph at top of figure plots Coherent Infomax values for each processor during learning). The final structure of weight strengths for each module reflected the edge variable (top of figure represents weight strengths using the same format of figure 7 in the target article). Similar results were also obtained with longer chains of processors, with different input features (lines against background), and also with partial overlap of receptive fields (25% of the total surface area) between adjacent processors.

3 Discovery of surface depth from multiple cues

Having shown that contextual signals could drive discovery of correlated functions even at long distances, a further set of experiments was run where a neural module attempted to drive another module to extract depth from stereo images. Depth perception from stereo images requires a non-linear transformation which could be learned by a multi-layer perceptron, but not with self-organized hebbian learning (unless some specific constraints are imposed on the network architecture). This seemed a suitable problem to test whether contextual signals arising from another neural module could guide learning of a computationally difficult function. Also, stereo-depth had already been used as a test problem within a similar framework [1] (see section 1.4, paragraph 4 of the target article for a discussion of important differences).

The neural network architecture was composed of two modules, each one being presented with input variables coherently varying across modules (see bottom of figure). One module, which was composed of two processors linked by reciprocal contextual connections, was presented with the same edge-contrast input described above. The other module, composed of a single multi-layer processor, was presented with two horizontal "slices of pixels", each one extracted from the images forming a random-dot stereogram. During learning, variation of the sign of the edge was correlated with a left or right shift in one of the random-dot images. The hidden unit and the output unit of the stereo-depth module received contextual signals from the output units of the edge-contrast network through unidirectional connections. Although these input patterns might be seen as multiple retinal cues correlated with depth (such as direction of contrast between foreground and background, and corresponding direction of binocular disparity), the choice of this particular setup was motivated mainly by coherence with previous experiments described in section 2 and by constraints ii) and iii) suggested in section 1. Several experiments were run where both modules simultaneously learned with a new set of initial random strengths for RF and CF connections. On average, the stereo-depth module successfully learned in 75% of the experiments to correctly signal the direction of retinal shift (the edge-contrast module always learned successfully), as it can be seen by the value of Coherent Infomax (graphs at bottom of figure). Several types of random-dot stereograms were used in these experiments, varying the size of the image, using a gaussian image-centered shift probability for each pixel, and adding various levels of uniform noise. Although these variations did not have an influence on whether or not the stereo-depth network learned successfully, the cases of failure took place when the edge-contrast network learned very quickly. As a control experiment, we observed that a stereo-depth module composed of two or more interconnected processors without contextual guidance from other modules could never discover the sign of the shift. As expected, the strength of the contextual connections to the hidden unit of the stereo-depth network was smaller than that of the connections to the output unit, because the hidden unit activation cannot be perfectly correlated with the direction of shift (if that was the case, a simple perceptron would be sufficient). Nonetheless, from a set of other experiments, it was clear that both contextual connections to the hidden unit and to the output unit were necessary for the stereo-depth network to learn correctly.

4 Conclusion

The experiments described here do not rule out the possibility that discovery of non-linear high-order coherent variables could simply arise from bottom-up information processing and reciprocal lateral interactions; rather, they intend to sugggest a further powerful functionality that unidirectional contextual connections might serve, that of an internal teacher. This possibility not only is a step forward in bridging the gap between a powerful learning paradigm and biological plausibility, but it also makes the theory suggested by the authors even more appealing as a common foundation of cortical computation.

Acknowledgments

Part of this work was supported by a Network grant from the Human Capital and Mobility Program of the European Community. Special thanks to Darragh Smyth for several interesting discussions.

References

[1] S. Becker and G. E. Hinton. A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355:161–163, 1992.

[2] F. Crick. The recent excitement about neural networks. *Nature*, 337:129–132, 1989.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Representations by Back-Propagation of Errors. *Nature*, 323:533–536, 1986.

[4] D. Zipser and R. A. Andersen. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331:679–684, 1988.

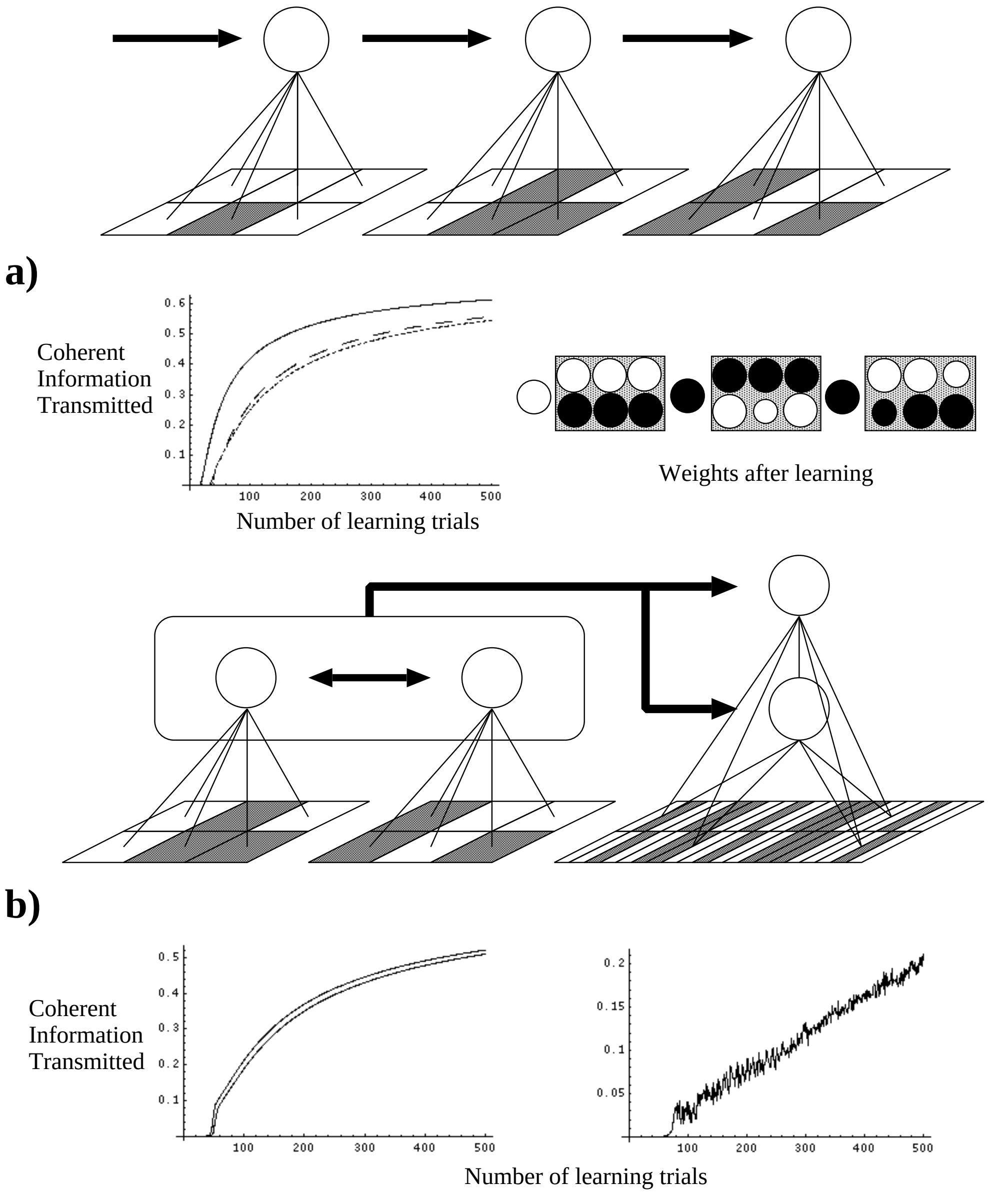

Figure 1: **a)** Top: contextually supervised edge-contrast detection (black=-1; white=+1). Thick arrows are contextual connections. Bottom: Coherent Information transmitted by each processor during learning (solid line: processor 1; dashed line: processor 2; dotted line: processor 3) and final CF and RF structure (black is inhibitory, white is excitatory, size is proportional to strength). **b)** Top: contextually supervised stereo-depth. Each output unit of the edge-contrast module sends a contextual signal to each unit of the stere-depth network (thick arrows). Input to stereo network consists of a right (top line) and left (bottom line) retinal image. Left image is shifted one pixel to the right. Bottom: Coherent Information transmitted by edge-contrast network (graph on the left) and by output unit of stereo-depth network (graph on the right).