



Report

Supervised and self-supervised learning with two sites of synaptic integration?

Author(s):

Körding, Konrad; König, Peter

Publication Date:

2001

Permanent Link:

<https://doi.org/10.3929/ethz-a-004264631> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Supervised and Self-Supervised Learning with two Sites of Synaptic Integration?

Konrad Paul Körding, Peter König,

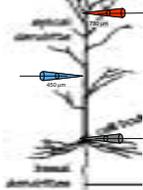
Institute of Neuroinformatics

ETH/UNI Zürich

Introduction

In many neural network models two independent variables are used to describe each neuron. One variable, called activity is the bottom-up information involved in signal processing. The performance of the network is typically measured by a goal function, e.g. minimum square error. Cells have a second variable, necessary for learning typically contains the derivative of this goal function with respect to the cell's input and thus allows the weight update. Since there did not seem to be a physiological basis for two variables the above abstract learning rules have often been considered biologically unrealistic. However recent research on the properties of cortical pyramidal cells shows that these cells have two sites of synaptic integration, the basal and the apical dendrite. Here we demonstrate that (1) with moderate assumptions neurons with two sites of synaptic integration can implement supervised as well as unsupervised learning mechanisms: the backpropagation of error algorithm (Werbos 1994/1974, Tesauro & Sejnowski 1990) and a self supervised coherence maximization algorithm; (2) that this is not directly possible with a cell model where just one integration is performed.

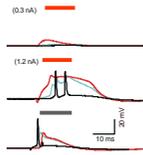
Two Sites of Synaptic Integration



The apical dendrite of layer V pyramidal neurons acts, in addition to the soma, as a second site of synaptic integration (Larkum et al 1999). The two sites interact in characterized ways:

- (1) to the apical dendrite by actively back-propagating dendritic action potentials.
- (2) to the soma via actively propagating slow regenerative calcium spikes that cause bursts.

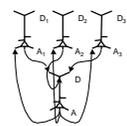
The cellular Physiology



- (1) Injection of sub-threshold current into the apical dendrite leads to strongly attenuated signals at the soma.
- (2) Injection of supra-threshold current into the apical dendrite leads to a calcium spike, a regenerative event and a postsynaptic burst of two spikes in this case
- (3) Injection into the soma leads to single spikes.

Following these experiments we describe each cell by two variables, somatic activity A and dendritic burst-rate D .

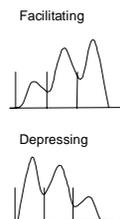
Afferents to Apical and Basal Dendrites



The apical dendrites receive local inhibitory projections, top down projections and long range cortico-cortical projections (Salin & Bullier 1995).

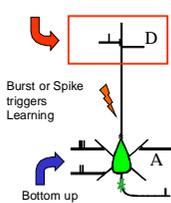
The basal dendrites of the considered cells receive direct subcortical afferents (e.g. the koniocellular pathway in visual cortex) and projections from layer 4 spiny stellate cells. These are the main recipients of afferents from sensory thalamus or from areas lower in the cortical hierarchy. Thus bottom-up input targets mainly the basal dendritic tree.

Dynamic Synapses



Synapses show temporal behaviors called facilitation and depression (Markram 1997). Facilitation has the effect that if two presynaptic events come within a short interval that the synaptic current is stronger. Depression has the effect that if a presynaptic cell is spiking fast the postsynaptic response is decreased. Highly facilitating synapses are burst detectors, depressing synapses single spikes detectors. Synapses can be characterized by the effect E of presynaptic activity A and presynaptic burst rate D : $E = VD_{pre} + WA_{pre}$, two signals are transmitted through the same axon.

Top down, Lateral



The Weight-update

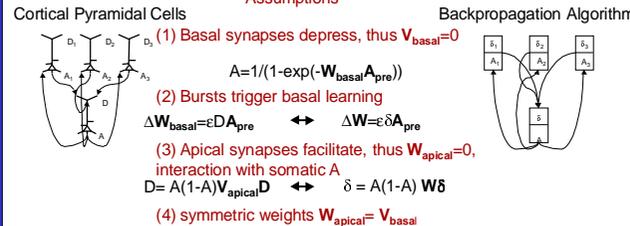
Experiments on hippocampal slices by Pike et al. (1999) support the idea that post-synaptic bursting can trigger LTP. Thus bursts and action potentials are potential triggers of synaptic plasticity.

Here we consider V and W and only changes proportional to the presynaptic A_{pre} and D_{pre} . This yields the following formula for the weight change:
 $\Delta W = f_1(A, D)(M_1 A_{pre} + M_2 D_{pre})$
 $\Delta V = f_2(A, D)(M_3 A_{pre} + M_4 D_{pre})$ where the M are mixing parameters which contain the rate of learning.

Backpropagation Learning

Could the described system do backpropagation learning? What assumptions are necessary?

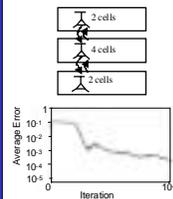
Assumptions



Biology could clearly implement backpropagation learning given the physiology. Presently there is no evidence for (1), (3) and (4) and we doubt that it is implemented that way in biology.

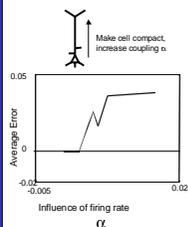
Learning the Xor Function

The backpropagation algorithm can approximate general nonlinear functions. Using a layer of sigmoid hidden neurons a perceptron can learn to approximate the Xor function, a popular benchmark system for backpropagation learning. The upper figure sketches the network used, the lower graph shows the residual error as a function of the iteration. As the weights are changed the network is doing gradient descent on the error function.



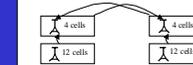
Do we need two Sites?

Neurons are often looked upon as being point systems, which can adequately be described by one variable. To test whether a comparable behavior could result with just one site of synaptic integration we changed the learning rule so that instead of ΔW being proportional to δ it is proportional to $(\delta + \alpha A)$. Learning thus is a mixture of backpropagation learning and hebbian learning. The left graph shows that the performance of the system is strongly impaired if the learning is a mixture of backpropagation and hebbian learning.



For correct learning of nonlinear functions with the backpropagation algorithm two sites of synaptic integration are necessary.

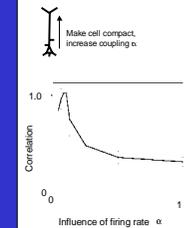
Self-supervised Learning



An alternative to learning with backpropagation alone is to use local goal functions (self supervision) that lead for example to the maximization of coherent information between separate streams (e.g. Becker & Hinton 1992).

We use $D = (WA_{pre} + \alpha A) = \max(WA_{pre} + \alpha A)$ and $A = G(A^{(i)}W_i - I_i)$, where G is the linear threshold function. Plotted in the graph is the normalized correlation, a measure of the performance of the network against the iteration. This graph shows that self supervised algorithms can also be implemented by the proposed mechanism with two sites of synaptic integration.

Do we need two Sites?



Once again we vary the mixing parameter of learning and processing α and measure the normalized correlation. The graph shows that increasing the effect of the postsynaptic activity on learning strongly affects the performance of the system. If α is small the cells learn according to the specified goal but mostly transmit information about non local stimulus properties (If α is very small the lateral weights fail to converge). If α is large the cells do not learn according to the set goal but represent local information.

For correct self supervised learning two sites of synaptic integration are necessary. One site leads to a tradeoff of local properties against quality of learning.

Discussion

With the given physiology backpropagation learning would be possible. The proposed mechanisms allow a smooth combination of supervised and unsupervised techniques. Backpropagation learning is extremely slow since only global properties of the stimulus are used for learning. Self supervised learning uses global context and local context in space, time and modality to guide learning. Thus self supervised learning can use more information for learning. Thus, especially when combined with supervised learning, we assume it to outperform purely supervised systems as research on self supervised networks reaches a level of refinement comparable to those of supervised systems. Instead of propagating back the global error through the whole system, locally generated goal functions can lead to local errors that propagated down a layer or two. Both backpropagation systems and self-supervised systems need two variables per cell. Using only one site of integration per cell severely distorts their ability to correctly reach their learning goals.

References

Becker, S., & Hinton, G.E. (1992). Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355, 161-163.

Larkum, M.E., Zhu, J.J., & Sakmann, B. (1999a). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398, 338-341.

Markram, H., Wang, Y., Tsodyks, M., Differential signaling via the same axon of neocortical pyramidal neurons. *Proc. Natl. Acad. Sci. USA*, 98 5323-5328

Pike, F.G., Meredith, R.M., Olding, A.W.A., & Paulsen O. (1999). Postsynaptic bursting is essential for 'Hebbian' induction of associative long-term potentiation at excitatory synapses in rat hippocampus. *J. Phys.*, 518.2, 571-576.

Salin, P.A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiol. Rev.*, 75, 107-154.

Tesauro, G & Sejnowski, T.J (1990) a parallel network that learns to play backgammon *Artificial Intelligence Journal* 39, 357-390

Werbos, P.J. (1974/1994) The roots of Backpropagation NY:John Wiley & Sons including the PhD thesis of Werbos P from 1974