



Novelty detection in a Kohonen-like network with a long-term depression learning rule

Dionyssios Theofilou*, Volker Steuber, Erik De Schutter

Laboratory of Theoretical Neurobiology, Born-Bunge Foundation, University of Antwerp-UIA, Universiteitsplein 1, Antwerp 2610, Belgium

Abstract

In the cerebellar cortex, long-term depression (LTD) of synapses between parallel fibers (PF) and Purkinje neurons can spread to neighboring ones, independently of their activation by PF input. This spread of non-specific LTD around the activated synapses resembles how units are affected in the neighborhood of the winner in a Kohonen Network (KN). However in a classic KN the weight vectors become more similar to the input vector with learning, while in the LTD case they should become more dissimilar. We devised a new LTD-KN where units, opposite to the classic KN, decrease their response (LTD-like) when a pattern is learned and we show that this LTD-KN functions as a novelty detector.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Cerebellum; Long-term depression; Self-organizing map; Kohonen network; Novelty detector

1. Introduction

Cerebellar Purkinje cells receive input from 150 000 parallel fibers, the axons of granule cells, and a single climbing fiber, the axon of an inferior olivary neuron. When a climbing fiber signal reaches the Purkinje cell repeatedly in conjunction with a signal from a parallel fiber, the synapse that received parallel fiber input when the climbing fiber signal arrived becomes depressed. According to the Marr–Albus–Ito theory [1,6,9] this mechanism may support motor learning, with the climbing fiber signal representing an error signal, and LTD leading to the depression of synapses responsible for the inaccuracy or miscalculation of the motor command.

* Corresponding author.

E-mail address: dion@bbf.uia.ac.be (D. Theofilou).

However this hypothesis has been challenged in the past [2] and seems to be in conflict with recent experimental results: First, parallel fiber stimulation alone can cause significant calcium influx and lead to LTD, without a climbing fiber signal present [4,3]. Second, the depression caused by the parallel fiber activity may not be specific to the synapses that receive the stimulus, but it may spread further to adjacent synapses that were not activated by the PF input [4,7,8].

According to the experimental findings, this non-specific LTD can be described by a Gaussian function with a half-width of about 50 μm around the activated synapses. It has been suggested that LTD may modify 600 times more inactive neighboring synapses than active synapses on a single Purkinje cell [8].

The way how LTD spreads to the neighboring synapses resembles the way neighboring units are influenced around the winner node in a Kohonen neural network [5]. During learning, the node with the strongest response when a pattern is presented to the Kohonen network is selected as the central winning output node. This node is altered as to maximize future response and increase the probability that it will be selected again when a similar input is presented. Moreover, it is not only the winner node which is affected during learning, but also all surrounding nodes, selected by a neighborhood function, usually a Gaussian. In contrast to the Kohonen network where the weight vectors of the output units become more similar to the input vector, cerebellar LTD should lead to a movement of weight vectors away from the input. In the following we will describe a modified version of a Kohonen network with an LTD learning rule.

2. The LTD-like learning rule

First, we normalize all input vectors before applying them to the network and the weight vectors at the initial phase and after every presentation of a pattern:

$$\sum_i (x_i)^2 = 1, \quad \sum_j (w'_{r'j})^2 = 1.$$

When a pattern is presented, the response of every node is given by the dot product of input vector and weight vector. The dot product is the measure of similarity between weight and input.

$$y_{r'} = \vec{x} \vec{w}_{r'}.$$

The node with the maximum response and therefore the smallest distance to the current input vector is the node r :

$$y_r = \max_{r'}(y_{r'}).$$

This node is chosen as the winner of the competition in the LTD-like network. Then the change of the weight vectors for winner and neighborhood-nodes is determined by an inverse and negative version of the classic Kohonen learning rule:

$$\Delta \vec{w}_{r'} = -a h_{rr'} g(\vec{x} - \vec{w}_{r'}),$$

where

$$g(\vec{x} - \vec{w}_{r'}) = \begin{cases} \frac{\vec{x} - \vec{w}_{r'}}{\|\vec{x} - \vec{w}_{r'}\| \delta_1} & \text{for } \|\vec{x} - \vec{w}_{r'}\| \leq \delta_1, \\ \frac{\vec{x} - \vec{w}_{r'}}{\|\vec{x} - \vec{w}_{r'}\|^2} & \text{for } \delta_1 < \|\vec{x} - \vec{w}_{r'}\| < \delta_2, \\ 0 & \text{for } \|\vec{x} - \vec{w}_{r'}\| \geq \delta_2, \end{cases}$$

δ_1, δ_2 are constants, points of maximum and minimum values of the g function, $h_{rr'}$ is a Gaussian neighborhood function centered at the winning unit r .

$$h_{rr'} = e^{-((r-r')^2/2\sigma^2)},$$

where σ defines the radius of the neighborhood which decreases over time and a is a positive learning rate that decreases toward zero as the learning progresses. The evolution of the weight vectors during learning is shown in Figs. 1 and 2.

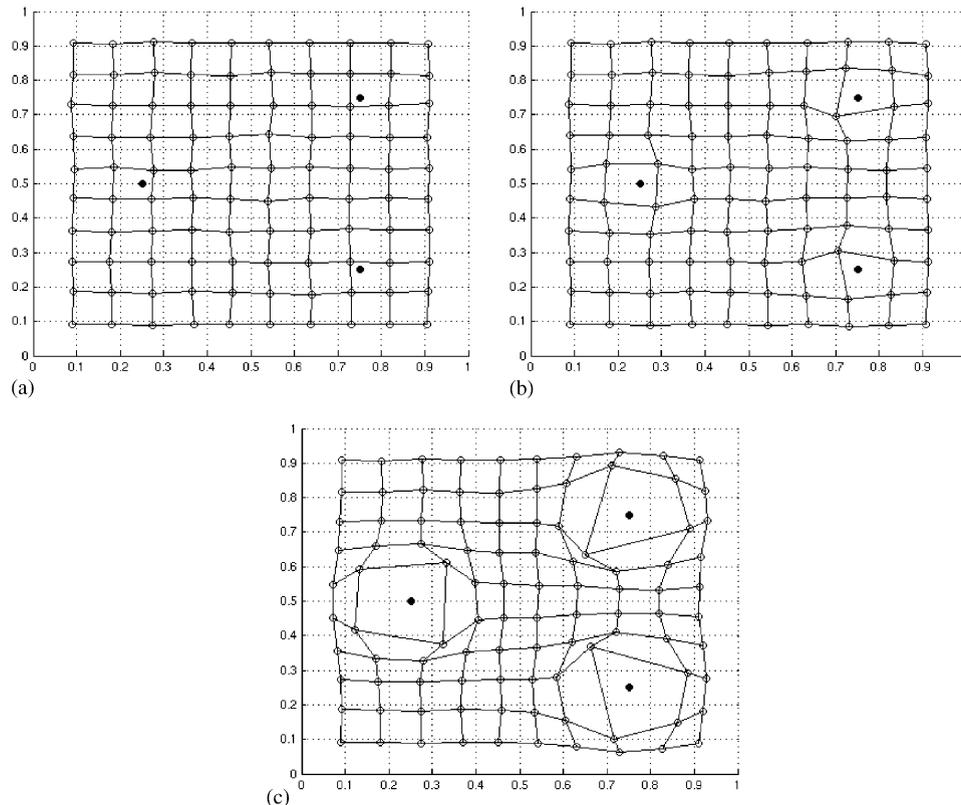


Fig. 1. A two-dimensional case of the network with the LTD-like learning rule is shown in this figure. Contrary to the classic Kohonen network where weight vectors move close to the input distribution, in the LTD-like case weights move away from it. This has as an effect of decreasing future activation with learning, corresponding to the synaptic depression effect in Purkinje cell synapses. Filled dots, represent the center of a Gaussian input distribution. Empty dots represent the weight vectors of the units in the Kohonen-like LTD network.

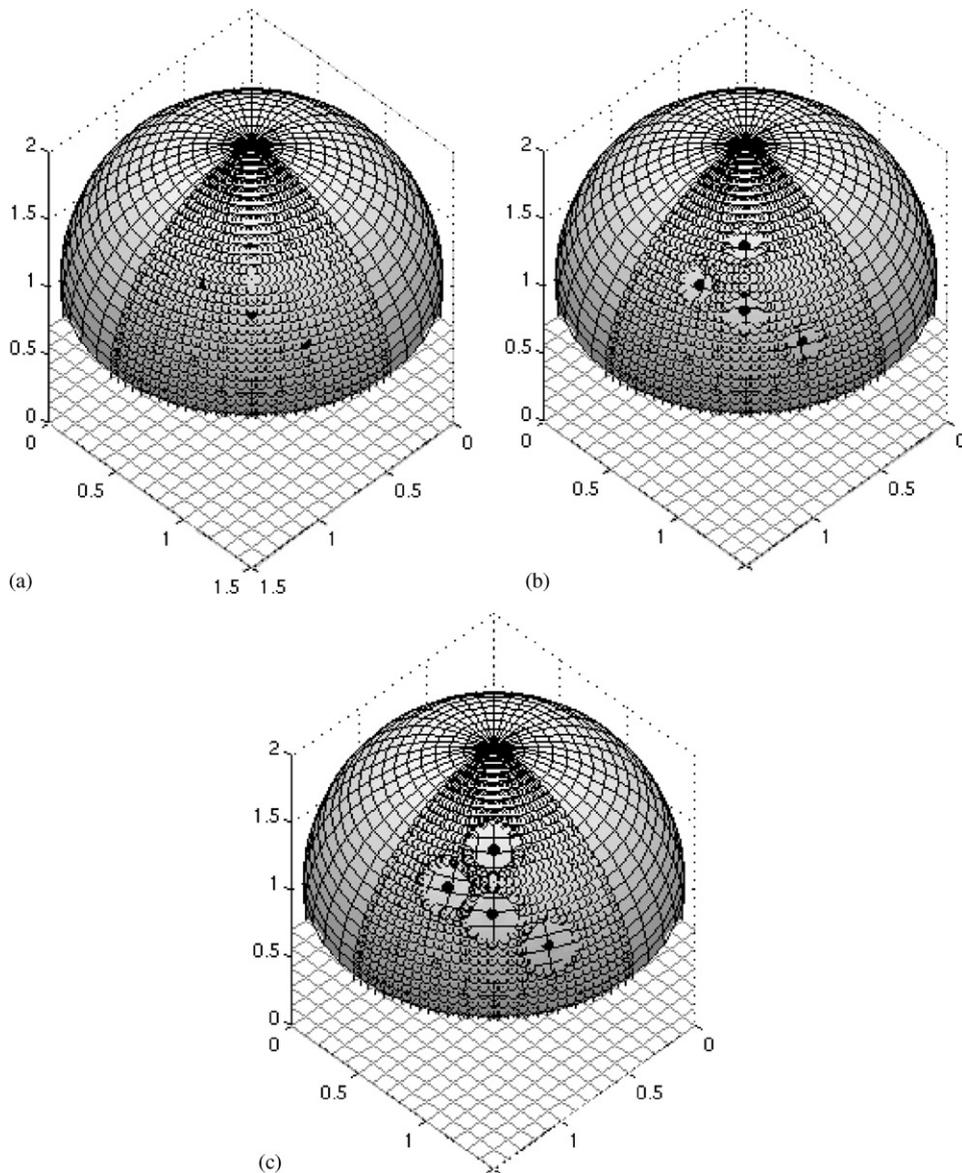


Fig. 2. The three-dimensional case of the network with the LTD-like learning rule is shown in this figure. The grid of empty dots represents the weight vectors, while the filled dots are the centers of input distributions. Additionally, all vectors are normalized (thus projected onto the surface of a unitary sphere) and comprise components in the range $[0, 1]$. On the left, the network before learning is shown, where all weights are randomly distributed. As learning progresses the weight vectors move away from the input vectors (center and right figure). With respect to the excitatory parallel fiber—Purkinje cell synapses, only positive weights are biologically plausible, thus only the first quadrant of the sphere contains weight units.

3. A novelty detector

As an effect of learning, all patterns used in the training set and all patterns similar to them give decreasing activation values as shown in Fig. 3. All patterns dissimilar to the training set result always in a stable high activation, both during and after learning. The differentiation of responses between patterns similar to training-patterns (known) and patterns not ever used in the training set (novel) comes eventually as a consequence of learning. Thus, an important property of the LTD Kohonen-like network is that it can function as a novelty detector. To demonstrate this we have made a map of all possible input patterns for the network (differing on small discrete step intervals) and show these results in Fig. 4. This activation map of possible inputs shows that the network can discriminate the known patterns used for learning and all patterns similar to them by providing a lower activation than for novel ones.

Novelty detection can be used to support both motor control and cognitive functions of the cerebellum. For example, the ability to make fast distinctions between familiar and novel patterns can support learning in many contexts. Moreover control systems can use novelty detection to distinguish between expected and unexpected patterns (even if the latter are not truly novel). Therefore this new interpretation of cerebellar learning could allow the cerebellar circuitry to perform many useful tasks fast.

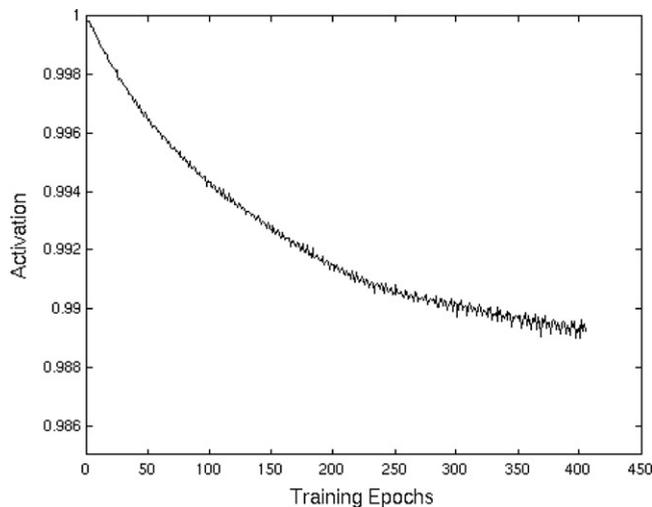


Fig. 3. The decrease of activation as learning progresses. The graph shows how the activation for a specific input similar to one of the learning patterns evolves during the first 400 training epochs.

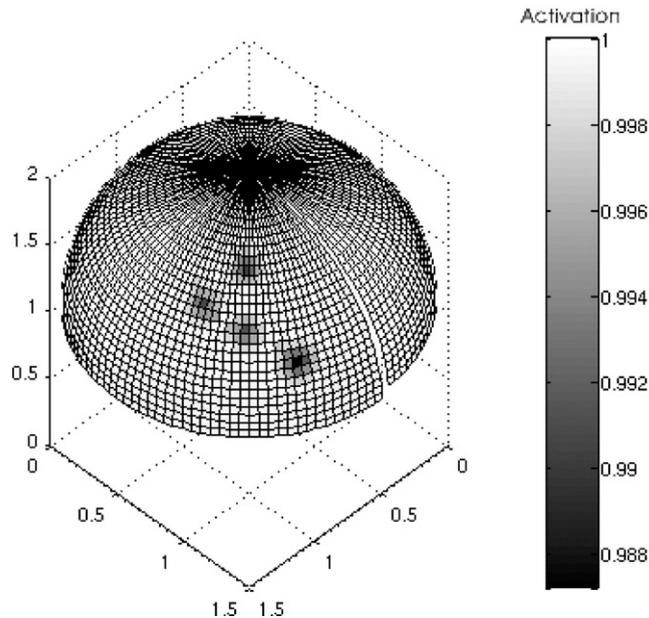


Fig. 4. In this figure all possible activation responses of the network for the input space after learning are shown. The input space contains darker areas which give lower activation and represent inputs similar to the ones used for training (learned patterns), while the areas with light color represent inputs not similar to the ones used in training (novel patterns). The network is a novelty detector for all possible patterns not similar to the ones used for training, by providing a higher activation response for them.

4. Conclusions

In the cerebellar cortex, long-term-depression of synapses between parallel fibers (PFs) and Purkinje cells is analogous to moving the synaptic weight vector away from the PF input vector. We have developed a Kohonen-like network with a LTD learning rule and shown that this network can function as a novelty detector. This novelty detection suggests new interpretations of cerebellar function in motor control and cognitive tasks

Acknowledgements

Thanks to Avrama Blackwell for her help and ideas when this project was initially considered at the EU computational neuroscience course in Trieste. This project was financially supported by the EC grant QLRT-2000-02256.

References

- [1] J.S. Albus, A theory of cerebellar function, *Math. Biosci.* 10 (1971) 25–61.

- [2] E. De Schutter, Cerebellar long-term depression might normalize excitation of Purkinje cells: a hypothesis, *Trends Neurosci.* 18 (1995) 291–295.
- [3] J. Eilers, H. Takechi, E.A. Finch, G.J. Augustine, A. Konnerth, Local dendritic Ca²⁺ signaling induces cerebellar long-term depression, *Learn. Mem.* 4 (1) (1997) 159–168.
- [4] N.A. Hartell, Strong activation of parallel fibers produces localized calcium transients and a form of LTD that spreads to distant synapses, *Neuron* 16 (1996) 601–610.
- [5] T. Kohonen, Self-organised formation of topologically correct feature maps, *Biol. Cybernet.* 43 (1982) 59–69.
- [6] D. Marr, A theory of cerebellar cortex, *J. Physiol.* 202 (1969) 437–470.
- [7] T. Reynolds, N.A. Hartell, An evaluation of the synapse specificity of long-term depression induced in rat cerebellar slice, *J. Physiol.* 527 (3) (2000) 563–577.
- [8] S.-H. Wang, L. Khiroug, G.J. Augustine, Quantification of spread of cerebellar long-term depression with chemical two-photon uncaging of glutamate, *Proc. Natl. Acad. Sci.* 97 (15) (2000) 8635–8640.
- [9] M. Ito, *The cerebellum and neural control*. Raven Press, New York, 1984.