Novelty detection in a Kohonen like network with a Long Term Depression learning rule

Dionyssios Theofilou, Volker Steuber and Erik De Schutter
Lab. for Theoretical Neurobiology, Born-Bunge foundation,
University of Antwerp (UIA), Antwerp – Belgium.

ABSTRACT

In the cerebellar cortex, long term depression (LTD) of synapses between parallel fibres (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 fuctions as a novelty detector.

SUMMARY

INTRODUCTION

Cerebellar Purkinje cells receive input from 150000 parallel fibres, the axons of granule cells, and a single climbing fibre, the axon of an inferior olivary neuron. When the signal of the climbing fibre reaches the Purkinje cell repeatedly in conjunction with a signal from a parallel fibre, the synapse that received parallel fibre input when the climbing fibre signal arrived becomes depressed. According to the Marr-Albus theory (Marr 1969, Albus 1971, Ito 1982) this mechanism may support motor learning, considering that the climbing fibre signal may represent an error signal, and LTD may lead to the depression of synapses responsible for the inaccuracy or miscalculation of the motor command.
However this hypothesis has been challenged in the past (De Schutter 1995)and seems to be in conflict with recent experimental results: First, parallel fibre stimulation alone can cause significant calcium influx and lead to LTD, without a climbing fibre signal present (Hartell 1996, Eilers et al 1997). Second, the depression caused by the parallel fibre 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 (Hartell 1996, Reynolds & Hartell 2000, Wang et al. 2000). The spread of LTD can be described by a Gaussian function with a half width of about 50 micrometers 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 (Wang et al. 2000).

The way that LTD spreads to the neighboring synapses resembles the way neighboring units are influenced around the winner node in a Kohonen neural network (Kohonen 1982). During learning, the patterns that are presented to the Kohonen network lead to the maximum response of a central winning output node. This node is altered as to maximize future response and increase the probability that it will be selected again in the future when a similar input is presented. But it is not only the winner node which is affected during learning, but also all nodes surrounding the winner, 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 results in the movement of weight vectors away from the input vectors. In the following we will describe a modified version of a Kohonen network with an LTD learning rule.

The LTD-like learning rule

We normalize all input vectors before applying them to the network and the weight vectors after every presentation of a pattern:

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

When a pattern is presented, the response of every node is given by the dot product of input vector and weight vector:

$$y_{r'} = \vec{x} \;\; \vec{w}$$

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 winner of the competition in the LTD-like network. 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:

$$g(\vec{x}-\vec{w}_{r'}) = \begin{cases} \dfrac{\vec{x}-\vec{w}_{r'}}{\|\vec{x}-\vec{w}_{r'}\|\delta_1} & , \; for \quad \|\vec{x}-\vec{w}_{r'}\| \le \delta_1 \\[2ex] \dfrac{\vec{x}-\vec{w}_{r'}}{\|\vec{x}-\vec{w}_{r'}\|^2} & , \; for \quad \delta_1 < \|\vec{x}-\vec{w}_{r'}\| < \delta_2 \\[2ex] 0 & , \; for \quad \|\vec{x}-\vec{w}_{r'}\| \ge \delta_2 \end{cases}$$

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

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

where σ defines the radius of the neighborhood which decreases over time.
a is a positive learning rate that decreases towards zero as the learning progresses.

RESULTS:

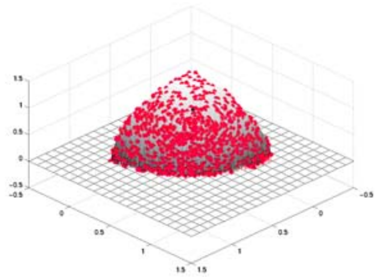

fig.1

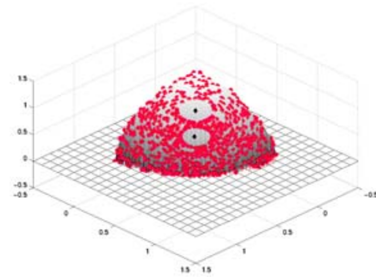

fig.2

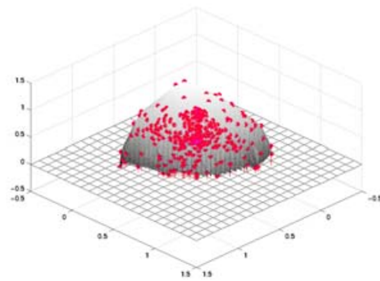

fig.3

The 3-dimensional case of the network with the LTD-like learning rule is shown in fig.1 and fig.2. The red points represent the weight vectors, while the black points are the input vectors. All vectors are normalized and comprise components in the range [0,1]. In fig.1, the network before learning is shown, where all weights are randomly distributed. As learning progresses the weight vectors move away from the input vectors (fig.2). In the case of the classic Kohonen algorithm, the weight vectors move towards the input vectors (fig.3).

NOVELTY DETECTION:
After the learning phase, an input vector which is similar to the patterns used for training will fall in an area the weight vectors have moved away from leading to a small response. Contrary, a novel input vector, different to the ones used for training, will be projected onto an area with a dense population of weight vectors, thus leading to a larger response of the LTD-like network. This guarantees that the network can function as a novelty detector, by responding maximally to patterns dissimilar to the ones used for learning.

BIBLIOGRAPHY

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

De Schutter, E. (1995) *Cerebellar long-term depression might normalize excitation of Purkinje cells: a hypothesis. Trends Neurosci.* **18**,291-295.

Eilers, J., Takechi H., Finch E.A., Augustine G.J. Konnerth A. (1997*) Local dendritic Ca2+ signaling induces cerebellar long-term depression*. Learn. Mem. 1997 **4-1**,159-168.

Hartell, N.A., (1996) *Strong Activation of Parallel Fibers Produces Localized Calcium Transients and a Form of LTD that Spreads to Distant Synapses.* Neuron **16**,601-610.

Kohonen, T. (1982). *Self-Organised Formation of Topologically Correct Feature Maps*. Biological Cybernetics **43**,59-69.

Marr, D., (1969) *A theory of cerebellar cortex*. J. Physiol. **202**,437-470.

Reynolds T., and Hartell (2000) *An evaluation of the synapse specificity of long-term depression induced in rat cerebellar slices.* J. Physiol., **527.3**,563-577.

Wang S.-h., Khiroug L., Augustine G.J., (2000*) Quantification of spread of cerebellar long-term depression with chemical two-photon uncaging of glutamate.* Proc.Natl.Acad.Sci. USA, **97-15**,8635-8640.