A biologically plausible implementation of error-backpropagation for classification tasks

Arjen van Ooyen[a], Pieter R. Roelfsema[b]

[a] *Netherlands Institute for Brain Research, Meibergdreef 33, 1105 AZ Amsterdam, The Netherlands*
[b] *Netherlands Ophthalmic Research Institute, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands*
A.van.Ooyen@nih.knaw.nl P.Roelfsema@ioi.knaw.nl

In: Kaynak, O., Alpaydin, E., Oja, E., Xu, L. (eds.) Artificial Neural Networks and Neural Information Processing - Supplementary Proceedings ICANN/ICONIP 2003, Istanbul, Turkey, June, pp. 442-444.

Error-backpropagation is a powerful method to train neural networks, but its current implementations lack biological realism. Here we present a novel scheme for implementing error-backpropagation in a biologically plausible way. Our scheme, called attention-gated reinforcement learning (AGREL), uses an "attentional" feedback signal to gate the plasticity of connections to hidden units. We show that the average changes in connection weights in AGREL are the same as the changes in weights in error-backpropagation.

We study a classification task with P input patterns that are classified into C mutually exclusive classes. The network that has to learn this classification is composed of N input units, M hidden units, and C output units (Fig. 1). The target output for input pattern p is $\mathbf{t}^p$, a vector in which all output units k have target activity 0 except the unit that encodes the target class for pattern p, unit $k = c_p$, which has target activity 1. Input patterns are applied to the input layer, and the activity of the hidden units, Y_j^p , is computed using the logistic activation function:

$$Y_j^p = \frac{1}{1 + \exp(-h_j^p)} \tag{1}$$

with

$$h_j^p = \sum_{i=0}^{N} v_{ij} X_i^p \tag{2}$$

where v_{0j} is the bias of unit j (X_0 always equals 1).

In the version of the backpropagation algorithm that is commonly used for classification problems with mutually exclusive classes [1], the activity of the output units, Z_k^p, is computed using the softmax activation function:

$$Z_k^p = \frac{\exp(a_k^p)}{\sum_{k'=1}^{C} \exp(a_{k'}^p)} \tag{3}$$

with

$$a_k^p = \sum_{j=0}^{M} w_{jk} Y_j^p \tag{4}$$

where w_{0k} is the bias of unit k (Y_0 always equals 1). The natural error function for these classification problems is the cross-entropy function for multiple classes [1], in which the error in the output for pattern p is defined as

$$Q_p = -\sum_{k=1}^{C} t_k^p \ln Z_k^p \tag{5}$$

where t_k^p is the target output for unit k, to be provided by a "teacher"; $t_{c_p}^p = 1$, and $t_k^p = 0$ for $k \neq c_p$. Connection weights are updated according to the gradient of the error surface. For the weights from the hidden to the output layer,

$$\begin{aligned} \Delta w_{jk} &= -\beta \frac{\partial Q_p}{\partial w_{jk}} = -\beta \frac{\partial a_k^p}{\partial w_{jk}} \frac{\partial Q_p}{\partial a_k^p} \\ &= -\beta Y_j^p \sum_{k'=1}^{C} \frac{\partial Q_p}{\partial Z_{k'}^p} \frac{\partial Z_{k'}^p}{\partial a_k^p} \\ &= \beta Y_j^p (t_k^p - Z_k^p) \end{aligned} \tag{6}$$

where β is a parameter that determines the learning rate. For the weights from the input to the hidden layer,

$$\begin{aligned} \Delta v_{ij} &= -\beta \frac{\partial Q_p}{\partial v_{ij}} = -\beta \frac{\partial h_j^p}{\partial v_{ij}} \frac{\partial Q_p}{\partial h_j^p} \\ &= -\beta X_i^p \sum_{k=1}^{C} \frac{\partial Q_p}{\partial a_k^p} \frac{\partial a_k^p}{\partial h_j^p} \\ &= \beta X_i^p Y_j^p (1 - Y_j^p) \sum_{k=1}^{C} (t_k^p - Z_k^p) w_{jk} \end{aligned} \tag{7}$$

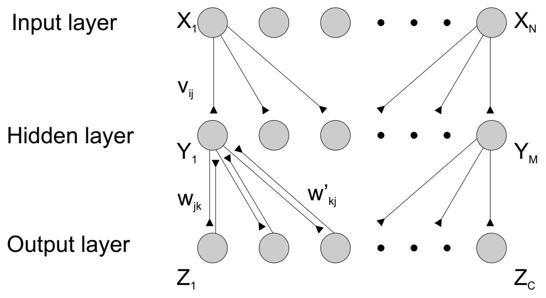

Figure 1: The three-layered network of AGREL. Feedback connections w'_{kj} from the output to the hidden layer gate the plasticity of the connections v_{ij} from the input to the hidden layer.

Updating according to these rules is biologically implausible, for two reasons. First, updating of the weights between the input and the hidden layer depends on w_{jk} and Z_k^p, information that is not locally available at the connection. Second, in each round of updating, a teacher has to provide the correct response t_k^p for each output unit.

In our new scheme, attention-gated reinforcement learning (AGREL), the activity in the output layer is determined by a winner-take-all rule. The winning output unit has activity 1, and the others are inactive. The softmax rule is used to determine the winning unit, whereby the total input a_k^p to a unit determines the probability that it will win the competition:

$$\Pr(Z_k^p = 1) = \frac{\exp(a_k^p)}{\sum_{k'=1}^{C} \exp(a_{k'}^p)} \quad (8)$$

with

$$a_k^p = \sum_{j=0}^{M} w_{jk} Y_j^p \quad (9)$$

In AGREL, as in other reinforcement learning schemes, evaluative feedback in the form of reward controls the changes in the connection weights. We assume, without loss of generality, that the amount r of reward received after correct classification equals 1, and that no reward is obtained in case of misclassification. With these choices, the average amount of reward expected with input pattern p is $\Pr(Z_{c_p}^p = 1)$, the probability that the correct output unit wins the competition. Suppose that, with input pattern p, output unit k has won the competition ($Z_k^p = 1$) and that this is the correct classification; i.e., $k = c_p$ and $r = 1$. The difference δ between the amount of reward obtained and the average amount of reward expected is then broadcasted to all the units of the network:

$$\delta = 1 - \Pr(Z_{c_p}^p = 1) \quad (10)$$

If another output unit, $k \neq c_p$, is selected, and therefore no reward is obtained, we take $\delta = -1$. The signal δ could be implemented as a diffusible messenger such as dopamine [2].

The change in the weights w_{jk} from the hidden to the output layer depends on Z_k^p, Y_j^p, and δ, information that is locally available:

$$\Delta w_{jk} = \beta Y_j^p Z_k^p f(\delta) \quad (11)$$

where β is a parameter that determines the learning rate and $f(\delta)$ is an expansive function that causes large weight changes for values of δ that are close to 1:

$$f(\delta) = \begin{cases} \delta/(1-\delta) & \delta \geq 0 \\ \delta & \delta < 0 \end{cases} \quad (12)$$

Thus, in each trial, only the connections to the output unit that has won the competition are updated, since for the other output units $Z_k^p = 0$.

By combining Eqs. (10) through (12), we can compute the average change in the weights w_{jk}. Pattern p is correctly classified with probability $\Pr(Z_{c_p}^p = 1)$, and the average change in the connection weights w_{jc_p} across trials equals

$$\begin{aligned} E(\Delta w_{jc_p}) &= \Pr(Z_{c_p}^p = 1)\beta Y_j^p \frac{\delta}{1-\delta} \\ &= \beta Y_j^p \left[t_{c_p}^p - \Pr(Z_{c_p}^p = 1)\right] \end{aligned} \quad (13)$$

An erroneous output unit $k \neq c_p$ is selected with probability $\Pr(Z_k^p = 1)$, and the average change in the connection weights w_{jk} is

$$\begin{aligned} E(\Delta w_{jk}) &= \Pr(Z_k^p = 1)\beta Y_j^p \delta \\ &= \beta Y_j^p \left[t_k^p - \Pr(Z_k^p = 1)\right] \end{aligned} \quad (14)$$

Comparison of Eqs. (13) and (14) with Eq. (6) shows that the average changes in weights in AGREL are the same as the changes in weights in error-backpropagation. Note that $\Pr(Z_k^p = 1)$ and $\Pr(Z_{c_p}^p = 1)$ in AGREL are equal to, respectively, Z_k^p and $Z_{c_p}^p$ in error-backpropagation.

The plasticity of the weights v_{ij} from the input to the hidden layer is gated by feedback of the winning unit through weights w'_{kj}, which approximate the weights w_{jk} of the feedforward connections. This feedback signal is physiologically plausible, since sensory neurons in the cortex receive attentional feedback when the object to which they respond is selected for action [3-5]. Changes in v_{ij} are determined by X_i^p, Y_j^p, δ, and the amount of feedback received from the output layer, all of which are locally available:

$$\Delta v_{ij} = \beta X_i^p Y_j^p (1 - Y_j^p) f(\delta) w'_{sj} \quad (15)$$

where w'_{sj} is the strength of the feedback connection from the winning unit s in the output layer to unit j in the hidden layer.

Each connection weight v_{ij} is updated both when input pattern p is correctly classified and when it is not. The average change in the connection weights v_{ij} across trials equals

$$\begin{aligned} E(\Delta v_{ij}) = \\ & \Pr(Z^p_{c_p} = 1)\beta X^p_i Y^p_j (1 - Y^p_j)\frac{\delta}{1-\delta} w'_{c_p j} \\ & - \sum_{k \neq c_p} \Pr(Z^p_k = 1)\beta X^p_i Y^p_j (1 - Y^p_j) w'_{kj} \\ = & \ \beta X^p_i Y^p_j (1 - Y^p_j) \sum_{k=1}^{C} \left[t^p_k - \Pr(Z^p_k = 1)\right] w'_{kj} \end{aligned} \quad (16)$$

Equation (16) is equivalent to Eq. (7). The plasticity of the feedback connections w'_{kj} is also governed by Eq. (11), and this maintains the equivalence of the weights of the feedforward and feedback connections.

In conclusion, we have shown that the error-backpropagation algorithm for classification tasks can be implemented as a reinforcement learning scheme. Various other schemes for implementing error-backpropagation in a more biologically plausible way have been proposed in the literature, including a second network for the error signals [6, 7], recirculation of activity in a recurrent network [8, 9], and two separate neuronal sites for the activity and the error signal [10]. However, in all these schemes, as in standard error-backpropagation, a biologically implausible teacher remains necessary to provide the correct output. In AGREL, as in other reinforcement learning schemes, it is not revealed how the network should have responded when the output of the network is wrong; the correct output is found by trial and error. AGREL is also superior to previous reinforcement learning schemes [11, 12], which are not as efficient as error-backpropagation in specifying how the connections to hidden units should be optimized. In AGREL, this so-called credit assignment problem is solved by the introduction of an attentional feedback signal. The combination of reinforcement learning and attentional feedback has yielded a learning scheme that changes connection weights just as standard error-backpropagation, but in a biologically plausible way.

References

[1] Bishop, C. M. *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, 1995).

[2] Pennartz, C. A. M. The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings. *Brain Res. Rev.* **21**, 219-245 (1995).

[3] Boch, R & Fischer, B. Saccadic reaction times and activation of the prelunate cortex: parallel observations in trained rhesus monkeys. *Exp. Brain Res.* **50**, 201-210 (1983).

[4] Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193-222 (1995).

[5] Moore, T. Shape representations and visual guidance of saccadic eye movements. *Science* **285**, 1914-1917 (1999).

[6] Zipser, D. & Rumelhart, D. E. Neurobiological significance of new learning models. In: E. Schwartz (ed.), *Computational Neuroscience*, MIT Press: Cambridge, MA, pp. 192-200 (1990).

[7] Tesauro, G. Neural models of classical conditioning: A theoretical viewpoint. In: S. J. Hanson, C. R. Olson (eds.), *Connectionist Modeling and Brain Function*, MIT Press: Cambridge, MA (1990).

[8] Hinton, G. E. & McClelland, J. L. Learning representations by recirculation. In: Anderson, D. Z. (ed.), *Neural Information Processing Systems*, American Institute of Physics, New York, pp. 358-366 (1988).

[9] O'Reilly, R. C. Biologically plausible error-driven learning using local activation differences: The generalised recirculation algorithm. *Neural Computation* **8**, 895-938 (1996).

[10] Körding, K. P. & König, P. Supervised and unsupervised learning with two sites of synaptic integration. *J. Comput. Neurosci.* **11**, 207-215 (2001).

[11] Barto, A. G. Learning by statistical cooperation of self-interested neuron-like computing elements. *Human Neurobiol.* **4**, 229-256 (1985).

[12] Mazzoni, P., Andersen, R. A. & Jordan, M. I. A more biologically plausible learning rule for neural networks. *Proc. Natl. Acad. Sci. USA* **88**, 4433-4437 (1991).