Matters temporal

Peter Dayan

Current evidence suggests that neural Hebbian learning in cortical and hippocampal synapses is fundamentally predictive rather than conventionally correlational. Much attention is focussing on what sort of predictions are acquired, and in what neural architectures. A recent paper by Rao and Sejnowski has suggested an interesting interpretation in terms of a popular predictive algorithm that has roots in psychology, computer science and engineering.

'*Prediction is difficult – especially about the future*', Yogi Berra noted. However, learning from the past to make predictions about the future is a critical computational component of many cognitive, perceptual and motor tasks, and so there is substantial interest in its neural underpinnings. Neural Hebbian learning rules have recently been realized to be even more 'Hebbian' than originally thought [1], in that they contain an asymmetry between the times of activation of pre- and post-synaptic cells. Such an asymmetry lies at the heart of any predictive algorithm, and there is substantial interest in the predictive capacities of temporally asymmetric Hebbian learning rules, and the relationship between such rules and other computational and engineering ideas about prediction learning [2], such as the method of temporal difference [3], which itself has rich psychological [4] and neural [5–7] roots. Rao and Sejnowski suggest a particularly concrete link between the two in their notable recent paper [8], arguing using a biophysically detailed model neuron, that if synapses were to implement a temporal-difference learning rule, then they would be expected to exhibit the sort of temporally asymmetric plasticity that has indeed been observed.

Here, I consider neural predictions from the perspective of temporal difference learning. What we will see is that although prediction is relatively straightforward at a systems level, it poses some interesting and tricky conceptual, architectural and mechanistic problems at the level of single neurons. Many of these problems were first discussed in a seminal paper on single cell prediction by Sutton and Barto [9], which is one of the main precursors to their later work on temporal difference learning [3,4].

Temporal difference learning was originally developed as such in the context of modeling classical conditioning [4], and this provides a convenient backdrop for our discussion. Consider a set of separate trials, in each of which a set of m stimuli is provided, the absence or presence of the i^{th} of which at time t is marked by $x_t(i) \in \{0,1\}$. Further suppose that a sequence of rewards is also provided, with r_t delivered at time t. Temporal difference learning solves the particular prediction problem, that Sutton and Barto suggested arises in classical conditioning, of learning weights $w(i)$ such that the stimulus at time t predicts the *sum* of future rewards in a trial. That is, the value $P_t = \sum_i w(i)\, x_t(i)$ should equal $\sum_{s>t}^{T} r_s$, where T is the last time in the trial. Since the target for prediction can be written in the form $\sum_{s>t}^{T} r_s = r_{t+1} + \sum_{s>t+1}^{T} r_s$, and P_{t+1} should equal $\sum_{s>t+1}^{T} r_s$, the successive predictions should be mutually consistent, with

$$P_t = r_{t+1} + P_{t+1}\,. \qquad \text{[Eqn 1]}$$

Temporal difference learning uses the difference between the two sides of this equation

$$\delta_t = r_{t+1} + P_{t+1} - P_t \qquad \text{[Eqn 2]}$$

as a prediction error associated with the stimuli $x_t(i)$, changing the weights as in the delta or equivalently Rescorla–Wagner [10] rule:

$$\begin{aligned} w(i) &\to w(i) + \varepsilon(r_{t+1} + P_{t+1} - P_t)\, x_t(i) \\ &= w(i) + \varepsilon\delta_t x_t(i) \qquad \text{[Eqn 3],} \end{aligned}$$

where ε is a learning rate. The involvement of P_{t+1}–P_t which, in a more temporally continuous setting would be a form of derivative of the prediction P_t, is what led to the name of the rule. Substantial theory is available as to circumstances under which this learning rule makes the predictions converge to the correct answers [11,12].

Learning rule 3 above has the characteristic that what might be thought of as the training signal from the prediction task (namely r_{t+1}) plays a subtly, but crucially, different role from that of the prediction P_t itself. That is, δ_t is *not* equal to a quantity $\delta'_t = (r_{t+1} + P_{t+1}) - (r_t + P_t)$ that we might call the 'activity difference'. Using the true temporal difference makes anatomical sense in a model of the dopamine system in primates, as there are plausibly separate discrete pathways conveying direct and predictive information about rewards, and in any event, it is only δ_t rather than P_t that is the required output for such purposes as learning actions that maximize rewards [13]. In fact, one can even see how the differentiating capacity of another of the recently prominent exotic findings about synapses, namely short-term depression [14–17], might provide a substrate for computing the temporal difference P_{t+1}–P_t. However, when P_t is taken to be the output (or, as Rao and Sejnowski suggest, a quantity more local to a synapse such as its membrane potential) of a single postsynaptic cell, it begs some interesting questions. How should we think of r_t as a privileged input to the cell? If a post-synaptic mechanism is responsible for computing P_{t+1}–P_t, then how can it avoid computing the incorrect activity difference δ'_t rather than δ_t? If it doesn't, and this is certainly the most natural conclusion from Rao and Sejnowski's interpretation of temporal difference learning leading to a temporally asymmetric Hebbian rule, then what are the consequences?

The answers to these questions, explored in Sutton and Barto's 1981 paper [9] (based on a slightly different formulation of a learning rule), find interesting resonance in various of Rao and Sejnowski's ingenious suggestions; Rao and Sejnowski also suggest some new ones. One set of ideas to maintain the temporal difference interpretation is to arrange the architecture, or the times of activation of the pre-synaptic units, or indeed the epochs in the trial over which plasticity pertains, so that the difference

1364-6613/02/$ – see front matter PII: S1364-6613(00)01851-9

between the activity difference δ'_t and the regular temporal difference δ_t never substantially affects the plasticity of a synapse. This happens, for instance, if there is only in fact one single reward right at the end of the trial [e.g. r_t=0, $t < T$; $r(T)$=1]. In a simplification of the version of this that Rao and Sejnowski consider, $r(T)$=1 comes from an action potential, caused by a privileged input to the postsynaptic cell, backpropagating up its dendritic tree. This makes positive the activity difference associated with presynaptic events initiated at time t=T–1 in a trial, thus engendering increases in synaptic efficacy. These, in turn, reduce the activity difference, by increasing P_{T-1}, until the difference reaches 0. This process can lead to the prior pre-synaptic events causing the post-synaptic cell to spike in a preditive manner. Rao and Sejnowski further suggest a specialized inhibitory connection architecture [18], which allows the predictive spike to cancel out the predicted spike (thus eliminating the effect of the difference between δ_t and δ'_t).

In the converse case, what happens if indeed δ'_t is used in the learning rule rather than δ_t? I don't know of compelling computational analyses of this case, other than the obvious point that the resulting learning rule looks like a correlational learning rule between the stimuli and the *differences* in successive outputs.

Rao and Sejnowski face the even trickier problem of making the learning rules work in the face of biophysically realistic timescales for synaptic currents and membrane potentials and the like. The most dangerous problem that arises is instability, that the learning rule can make the synaptic efficacies rise without bound. This happens when the biophysical mechanism for propagating information around the post-synaptic cell (backpropagating action potentials) lasts over a longer time scale than that involved in the derivative P_{t+1}–P_t. That can make the learning rule operate more like a regular correlational learning rule, and these are notoriously unstable. Synaptic saturation is suggested as a possible fix, although one might worry about a consequent loss of synaptic selectivity.

Altogether, the notion that temporally asymmetric Hebbian learning rules are best seen in predictive rather than correlational terms has been taken in various interesting directions. Rao and Sejnowski usefully add to our armoury of ways of approaching such rules, and remind us of an essential Yogic truth.

Acknowledgements

I am grateful to Raj Rao and Chris Watkins for helpful comments.
The author's research is funded by the Gatsby Charitable Foundation.

References

1 Sejnowski, T.J. (1999) The book of Hebb. *Neuron* 24, 773–776
2 Abbott, L.F. and Blum, K.I. (1996) Functional significance of long-term potentiation for sequence learning and prediction. *Cereb. Cortex* 6, 406–416
3 Sutton, R.S. (1988) Learning to predict by the methods of temporal difference. *Mach. Learn.* 3, 9–44
4 Sutton, R.S. and Barto, A.G. (1990) Time-derivative models of Pavlovian conditioning. In *Learning and Computational Neuroscience*, (Gabriel, M. and Moore, J.W., eds), pp. 497–537, MIT Press
5 Montague, P.R. *et al.* (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947
6 Schultz, W. *et al.* (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599
7 Schultz, W. (1998) Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27
8 Rao, R.P.N. and Sejnowski, T.K. (2001) Spike-timing dependent Hebbian plasticity as temporal difference learning. *Neural Comput.* 13, 2221–2237
9 Sutton, R.S. and Barto, A.G. (1981) Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135–170
10 Rescorla, R.A. and Wagner, A.R. (1972) A theory of Pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement. In *Classical Conditioning Vol. II: Current Research and Theory* (Black, A.H. and Prokasy, W.F., eds), pp. 64–69, Appleton-Century-Crofts
11 Sutton, R.S. and Barto, A.G. (1998) *Reinforcement Learning*, MIT Press
12 Bertsekas, D.P. and Tsitsiklis, J.N. (1996) *Neuro-Dynamic Programming*, Athena Scientific
13 Barto, A.G. *et al.* (1990) Learning and sequential decision making. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, (Gabriel, M. and Moore, J.W., eds), pp. 539–602, MIT Press
14 Grossberg, S. and Schmajuk, N.A. (1987) Neural dynamics of attentionally modulated Pavlovian conditioning: conditioned reinforcement, inhibition, and opponent processing. *Psychobiology* 15, 195–240
15 Abbott, L.F. *et al.* (1997) Synaptic depression and cortical gain control. *Science* 275, 220–224
16 Tsodyks, M.V. and Markram, H. (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl. Acad. Sci. U. S. A.* 94, 719–723
17 Thomson, A.M. (2000) Facilitation, augmentation and potentiation at central synapses. *Trends Neurosci.* 23, 305–312
18 Rao, R.P.N. and Sejnowski, T.J. (2000) Predictive sequence learning in recurrent neocortical circuits. In *Advances in Neural Information Processing Systems* (Vol. 12) (Solla, S.A. *et al.*, eds), pp. 164–170, MIT Press

Peter Dayan
Gatsby Computational Neuroscience Unit, University College London, UK WC1E 6BT.
e-mail: dayan@gatsby.ucl.ac.uk

Debate

Neuroecology and psychological modularity

Jonathan I. Flombaum, Laurie R. Santos and Marc D. Hauser

In a recent review, Bolhuis and Macphail challenge the thesis that specialized systems mediate the learning, encoding and retrieval of different types of information – what they call a neuroecological approach to learning and memory [1]. In particular, they challenge 'the arbitrary assumption that different "problems" engage different memory mechanisms' (p. 426), and the idea that this fact can be used to motivate neurobiological studies. To substantiate their claims, they appeal to data dealing with the neural substrates of song learning and food storage in birds. Recently, Hampton *et al.* [2] pointed out how Bolhuis and Macphail misrepresent these data and set-up a 'straw neuroecologist' with respect to the functionalist/adaptionist perspective. Here, we take up a different problem.

1364-6613/02/$ – see front matter PII: S1364-6613(02)01872-7