NEUROSCIENCE AND MACHINE LEARNING

Bursting potentiates the neuro–AI connection

For decades, researchers have wondered whether algorithms used by artificial neural networks might be implemented by biological networks. Payeur et al. have strengthened the connection between neuroscience and artificial intelligence by showing that biologically plausible mechanisms can approximate key features of an essential artificial intelligence learning algorithm.

Weinan Sun, Xinyu Zhao and Nelson Spruston

For many researchers, the connection between neuroscience and artificial intelligence (AI) is a natural one. Over the past several years, artificial neural networks, or ANNs—so-named because of their resemblance to brains—have been trained to perform specific tasks with superhuman performance. For example, artificial systems now routinely beat the best human players at complex games such as Go[1] and Dota[2], and they can solve enormously complex problems such as protein folding[3]. Neuroscientists therefore wonder whether these sophisticated systems can teach us something about the mechanisms of cognition in humans and other animals. On the other hand, AI systems are, for the most part, highly specialized. A system designed and trained to play games cannot fold proteins, and vice versa. In the domain of general intelligence, humans still outperform machines. Thus, practitioners of AI wonder whether neuroscience can offer clues as to how to achieve more general artificial intelligence.

In contrast with this sense of optimism about the neuro–AI intersection, however, many skeptics have argued that ANNs depend on computational methods that are so dramatically different from biological brains that the intersection is a mirage. Central to this argument is the key role played in AI systems by the error backpropagation algorithm[4], commonly referred to as 'backprop'. This mathematical algorithm is perceived by many as biologically implausible, for a variety of reasons.

A new paper in *Nature Neuroscience* by Payeur et al.[5] aims to overturn that argument, by showing that a collection of biologically feasible mechanisms can be combined to reproduce the essential features of backprop, thus allowing effective learning in hierarchical biological networks. Crucially, they demonstrate that implementing these mechanisms can achieve performance comparable to backprop-trained networks with similar architecture, an admirable feat for a biologically plausible network.

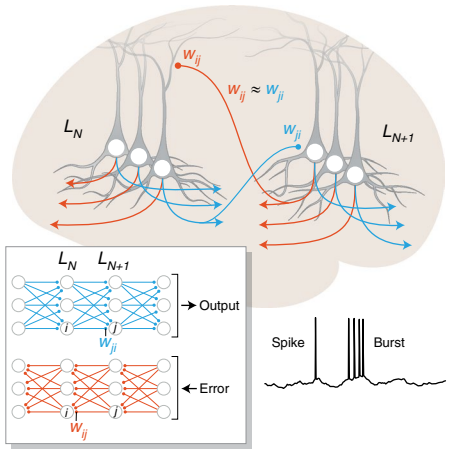

Fig. 1 | Does the brain solve the credit assignment problem using learning algorithms akin to back-propagation?. Artificial neural networks (bottom left) often rely on error backpropagation (backprop) to solve the credit assignment problem necessary for weight adjustment. The network performs a feedforward computation to determine an output, which is then compared to a target to determine an error. On a separate pass, the error is propagated through the network to adjust synaptic weights in the direction of improved network performance. Feedforward and feedback weights are symmetrical. In the brain (top), feedforward inputs from lower cortical areas to higher cortical areas and feedback from higher to lower areas both exist and may therefore carry the signals necessary for feedforward computation and feedback error propagation. Payeur et al. combine a collection of biologically plausible mechanisms by which these networks can: (1) use feedback (red) to steer the sign and magnitude of plasticity, (2) multiplex feedforward processing (blue) and feedback error propagation (red), and (3) promote symmetry in the feedforward and feedback weights ($w_{ji} \approx w_{ij}$) through learning. Together, this solves the credit assignment problem in hierarchical networks in a biologically plausible manner. Central to these mechanisms are feedforward projections from lower cortical areas (L_N) to higher cortical areas (L_{N+1}) and reciprocal feedback connections (L_{N+1} to L_N), the separation of feedforward and feedback synapses on distinct dendritic compartments, the promotion of bursting (bottom right) by feedback inputs, burst-dependent plasticity rules, and distinct short-term plasticity at feedforward vs feedback synapses. Illustrations by Julia Kuhl. Bottom left panel illustrating backprop is adapted from ref. [6], Nature Publishing Group.

Standard ANNs consist of multiple layers: an input layer, an output layer, and one or more 'hidden layers' in between. Each layer consists of a number of units, often compared to neurons, and each of these units is connected to all of the units in the next layer by adjustable weights, often compared to synaptic weights (Fig. 1). Inputs to the system sequentially activate layers of the network to produce a final output, which is then compared to the target the network is intended to reproduce.

ANNs can approximate very complex functions by adjusting the connection weights until performance error gets very small. Here, backprop plays a key role, as it ensures that weights are adjusted in a manner that promotes better performance, and since backprop ensures that synaptic changes follow the error gradient, learning is also efficient[6].

Successful learning requires that every weight in the network is adjusted to correct for its proportional contribution to the

error in the final output layer. This is called the 'credit assignment' problem. In ANNs implementing backprop, it is solved mathematically using the chain rule for derivatives. It is unclear, however, whether brains solve the problem in a similar way, and if they do, which mechanisms could achieve this. How could plasticity mechanisms at a synapse deep in the brain possibly know how much it had contributed to the error in the output of a massively complex computation across many layers of synaptic connections?

A natural way to solve such credit assignment problem biologically would be for an error signal to be transmitted synaptically in the network. To do so, pairs of neurons in adjacent layers would be connected reciprocally, via feedforward and feedback connections (Fig. 1). This kind of circuit arrangement is reasonable, but two further challenges confront the execution of learning via such a circuit. First, in ANNs, input–output computations and error-corrective backprop are computed on separate passes through the network, whereas in biological systems, it is unlikely such separate passes exist, so the two processes would need to be multiplexed in time. Second, backprop requires symmetrical weights between reciprocally connected neurons (Fig. 1). Payeur et al.[5] show that both problems can, in principle, be solved by biologically plausible mechanisms.

A few critical ideas, extracted from the literature on dendritic integration and synaptic plasticity, allowed Payeur et al.[5] to present a solution (Fig. 1). First, feedforward synapses are located on dendrites closer to the cell body, while feedback inputs are located on distal apical dendrites, much further from the soma. Second, feedback synapses on distal dendrites have been shown to cause dendritic spikes, which promote action potential bursting, especially when paired with spikes driven by perisomatic feedforward inputs[7]. Third, action potential bursting has been shown to tip the balance from long-term depression (LTD) of synaptic weights toward long-term potentiation (LTP)[8,9]. Combining these observations, the authors arrive at the conclusion that for any given neuron, the feedforward synaptic weight could be steered by the post-synaptic integration of feedforward and feedback inputs, with the feedback inputs transmitting the appropriate error signal. In addition, the authors propose a sliding LTP/LTD threshold, determined by the running average of burst probability, to maintain the stability of weight changes.

The multiplexing problem poses a challenge because biological networks must nearly simultaneously compute the feedforward input–output computation and the feedback-based error correction without interference. Payeur et al. show that another set of biologically plausible mechanisms, complementing the plasticity rules described above, can be combined to allow burst-driven plasticity to occur without influencing the feedforward network computation. Central to these mechanisms are distinct forms of short-term synaptic plasticity (depression and facilitation) for feedforward and feedback pathways, including effects on both excitatory and inhibitory neurons.

The weight symmetry problem is thorny, because it is typically solved in ANNs by copying feedforward weights to their feedback counterparts. This 'weight transport' problem[10] seems biologically implausible. Previous theoretical studies have shown that weight symmetry can be achieved in a variety of ways[11,12]. Payeur et al. build on these previous insights to show that weight symmetry can be achieved biologically using plasticity rules that depend on distinct burst-dependent plasticity rules for feedforward and feedback inputs. Both forms of plasticity, used to adjust the feedforward and feedback weights (w_{ji} and w_{ij}, respectively, in Fig. 1), are dependent on the event rate (spikes and bursts) in the lower-layer neuron (L_N in Fig. 1) and the burst rate in the higher-layer neuron (L_{N+1} in Fig. 1).

For the most part, the methods used by Payeur et al. are variants of methods that have been described previously. The central contribution of their work is to elegantly combine these methods in a single network model. To demonstrate that this combination of mechanisms is effective, they show that their model can compute the XOR operation and that it performs well on benchmark AI problems such as CIFAR-10 and ImageNet.

Many neuroscientists will find good reasons to question whether the methods put forward by Payeur et al. are actually used for credit assignment in biological brains. In our view, however, the work is fundamentally important because it shows that biological mechanisms can be combined to approximate the error backpropagation algorithm, thus challenging the perception that it is biologically implausible.

Going forward, the work provides a roadmap for scientists interested in the interface between neuroscience and AI. For theorists: are there other network architectures and biological mechanisms that can rival or surpass the performance of the network described by Payeur et al.? Notably, their deep model included convolutional layers with standard backprop, so it will be interesting to see how fully biological deep networks perform. For experimentalists: are plasticity rules such as those described here actually implemented in the biological circuits during learning? Are symmetrical synaptic weights, such as those observed previously in local circuits[13,14], also observed in the long-range projections thought to reflect feedforward and feedback components of multilayer circuits? These are just some of the experimentally tractable questions that are likely to motivate future research in this area.

ANNs, which are at the center of current AI systems, have borrowed key ideas from biological brains: for example, distributed representations and adjustable connections. On the other hand, neuroscience can benefit substantially from AI research stimulating new hypotheses about how the brain can potentially solve intelligent tasks[15]. Achieving a positive feedback loop to accelerate advancements in neuroscience and AI requires healthy debate and collaboration between experts in both fields. In the meantime, the study by Payeur et al. exemplifies that the neuro–AI interface is promising territory for those interested in intelligent animals and machines. ❐

Weinan Sun[1,2], Xinyu Zhao[1,2] and Nelson Spruston[1]✉

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA. [2]These authors contributed equally: Weinan Sun, Xinyu Zhao.
✉*e-mail: sprustonn@janelia.hhmi.org*

Published online: 13 May 2021
https://doi.org/10.1038/s41593-021-00844-2

References

1. Silver, D. et al. *Nature* **529**, 484–489 (2016).
2. OpenAI et al. Preprint at *arXiv* https://arxiv.org/abs/1912.06680 (2019).
3. Senior, A. W. et al. *Nature* **577**, 706–710 (2020).
4. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. *Nature* **323**, 533–536 (1986).
5. Payeur et al. *Nat. Neurosci.* https://doi.org/10.1038/s41593-021-00857-x (2021).
6. Lillicrap, T. P., Santoro, A., Marris, L., Akerman, C. J. & Hinton, G. *Nat. Rev. Neurosci.* **21**, 335–346 (2020).
7. Larkum, M. E., Zhu, J. J. & Sakmann, B. *Nature* **398**, 338–341 (1999).
8. Letzkus, J. J., Kampa, B. M. & Stuart, G. J. *J. Neurosci.* **26**, 10420–10429 (2006).
9. Sjöström, P. J., Turrigiano, G. G. & Nelson, S. B. *Neuron* **32**, 1149–1164 (2001).
10. Grossberg, S. *Cog. Sci.* **11**, 23–63 (1987).
11. Kolen, J. F. & Pollack, J. B. in *Proc. 1994 IEEE International Conference on Neural Networks (ICNN'94)* https://doi.org/10.1109/ICNN.1994.374486 (1994).
12. Akrout, M., Wilson, C., Humphreys, P. C., Lillicrap, T. & Tweed, D. *Adv. Neural Inf. Process. Syst.* {https://proceedings.neurips.cc/paper/2019/file/f387624df552cea2f369918c5e1e12bc-Paper.pdf (2020).
13. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. *PLOS Biol.* **3**, e68 (2005).
14. Cossell, L. et al. *Nature* **518**, 399–403 (2015).
15. Richards, B. A. et al. *Nat. Neurosci.* **22**, 1761–1770 (2019).

Competing interests

The authors declare no competing interests.