Enhancing Learning in Sparse Neural Networks: A Hebbian Learning Approach*

Alexander de Ranitz[1], Ardion Beldad[2], and Elena Mocanu[3]

[1] ATLAS, University of Twente, the Netherlands
[2] University College Twente, University of Twente, the Netherlands
[3] EEMCS Faculty, University of Twente, the Netherlands

Abstract. Artificial neural networks have proven to be capable of mastering many complex tasks. However, training such networks can be extremely resource intensive. In this research, the learning rule of a neural network trained using sparse evolutionary training (SET) is extended based on Hebbian theory. A mathematical formulation of Hebbian theory, encompassing inhibitory neurons and tailored for artificial neural networks, is proposed. The resulting novel algorithm, referred to as HebbSET, exhibits enhanced performance in terms of learning speed and final accuracy on two datasets. These findings underscore the potential of incorporating neuroscientific theories to enhance the capabilities of ANNs and bridge the gap between neuroscience and AI.

Keywords: Artificial neural networks · Hebbian theory · Sparse training

1 Introduction

In recent years, artificial neural networks (ANNs) have shown to be extremely powerful. However, most state-of-the-art models are extremely large and use a fully-connected architecture, requiring tremendous computational power and energy to train and run. Furthermore, research has shown that trained ANNs tend to have a weight distribution centred around 0, indicating that a significant number of connections is not meaningfully contributing to the output of the network [2]. Sparse neural networks, in which each neuron is connected to only a subset of the neurons in the following layer, can be used to reduce the number of (possibly unnecessary) connections, reducing computational load while performance stays approximately equal [1].

There are many differences between ANNs and biological neural networks. One such difference is the way they learn. A famous concept in neuroscience is Hebb's postulate, often summarised as: *"Neurons that fire together, wire together"* [3]. While this concept is fundamental to our understanding of how the human brain learns, modern ANNs that learn using gradient descent methods do not (explicitly) incorporate this idea. Here, it will be investigated whether Hebbian learning can be used to enhance learning in truly sparse neural networks trained using Sparse Evolutionary Training (SET) [4].

* This Extended Abstract is based on the full Bachelor Thesis of Alexander de Ranitz, defended in July 2023, which is available at the following link.

2 Hebbian Learning for Artificial Neural Networks

In order to be used to train ANNs, Hebb's postulate is extended and appropriately quantified. The intended behaviour of Hebb's postulate in ANNs can be summarised as follows: (1) If presynaptic neuron (i.e. the neuron generating the output) A is active and postsynaptic (i.e. the neuron receiving the input) neuron B is active, A's efficiency, as one of the neurons firing B, should be increased (i.e. the weight between A and B should increase in ANNs), and (2) If presynaptic neuron A is active and postsynaptic neuron B is inactive (suppressed), A's efficiency, as one of the neurons suppressing B, should be increased (i.e. the weight between A and B should decrease in ANNs). These two requirements are fulfilled by the following equation

$$\Delta w_{ij}^{l} = a_j^{l-1} \cdot (a_i^l - \overline{a_i^l}) \tag{1}$$

in which $\overline{a_i^l}$ represents the average activation of the postsynaptic neuron over a given time period. Equation 1 is combined with gradient descent to form the complete learning rule: $\Delta w_{ij}^{l} = -\alpha \cdot \dfrac{\partial J}{\partial w_{ij}^{l}} + a_j^{l-1} \cdot (a_i^l - \overline{a_i^l}) \cdot \lambda(t)$ where α and $\lambda(t)$ are (time-dependent) hyperparameters to control the learning rate and the relative strength of the Hebbian learning factor. It is hypothesised that incorporating this Hebbian learning term in the learning rule of SET can improve the performance of neural networks by strengthening important weights and reducing unimportant weights. This novel algorithm is named HebbSET.

3 Results & Conclusion

The results of the baseline SET algorithm and the HebbSET algorithm on the MNIST and Lung datasets are presented in Fig. 1.

Compared to SET, HebbSET displays an increased learning speed and a higher final accuracy. These differences are especially noticeable on the MNIST dataset. Further analysis of the resulting networks showed that the networks trained using HebbSET more effectively adapted to the input data and converged to a (locally) optimal topology more rapidly.

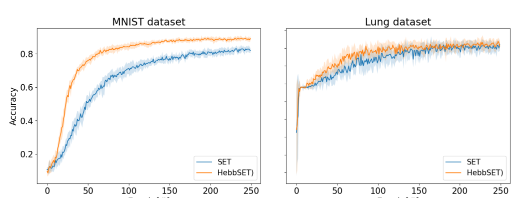

Fig. 1. Accuracy over time for SET and HebbSET on the MNIST and Lung dataset. Results are averaged over ten iterations. The shaded area shows the standard deviation.

In conclusion, incorporating elements of Hebbian learning in the training of ANNs can yield promising results, allowing ANNs to learn in a manner that is closer to what is believed to be happening in the human brain whilst simultaneously improving network performance.

References

1. Frankle, J., Carbin, M.: The lottery ticket hypothesis: Finding sparse, trainable neural networks (2019), https://doi.org/10.48550/arXiv.1803.03635
2. Han, S., Pool, J., Tran, J., Dally, W.J.: Learning both weights and connections for efficient neural networks (2015), https://doi.org/10.48550/arXiv.1506.02626
3. Hebb, D.O.: The organisation of behaviour: A neuropsychological theory. John Wiley and Sons (1949)
4. Mocanu, D.C., Mocanu, E., Stone, P., Nguyen, P.H., Gibescu, M., Liotta, A.: Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. Nature Communications **9** (2018). https://doi.org/10.1038/s41467-018-04316-3