Sparse neural networks

Russel Shawn Dsouza

Electronics and Communications Engg.
National Institute of Technology Karnataka
Surathkal, India - 575025

February 28, 2020

Overview

Motivation for sparsity: Network Theory

Fewer links than the maximum possible links
Ex: Social and computer networks

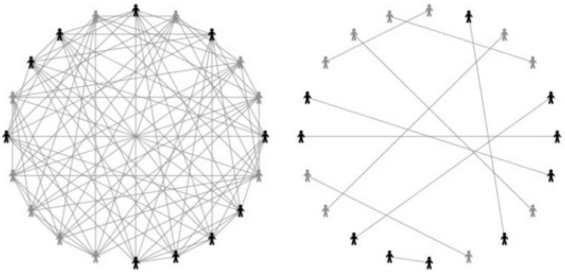

Figure: Dense vs Sparse social networks[1]

[1] S. Barnes, *Researchers Propose Social Network Modeling to Fight Hospital Infections*, https://umdrightnow.umd.edu/news/researchers-propose-social-network-modeling-fight-hospital-infections, [Online; accessed 26/02/2020], 2013.

Motivation for sparsity: Deep Learning (-1998)

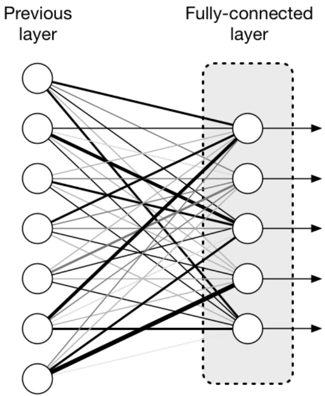

Figure: Fully connected networks[2]

[2] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," *Neural networks*, vol. 1, no. 2, pp. 119–130, 1988.

Motivation for sparsity: Deep Learning (-2019)

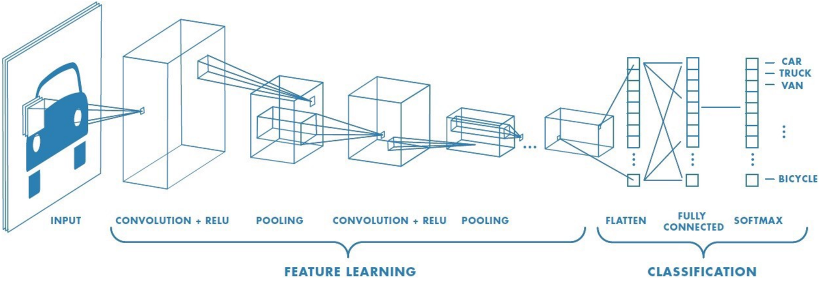

Figure: Convolutional neural network[3]

[3] S. Saha, *A comprehensive guide to convolutional neural networks: The ELI5 way*, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53, [Online; accessed 26/02/2020], 2018.

Motivation for sparsity: Neocortex

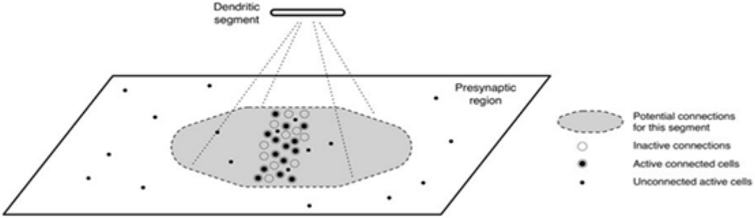

Figure: Sparse coding in the brain[4]

[4] numenta.com, *Sparse distributed representations*, https://numenta.com/neuroscience-research/sparse-distributed-representations/, [Online; accessed 26/02/2020], 2018.

Sparsely connected layers

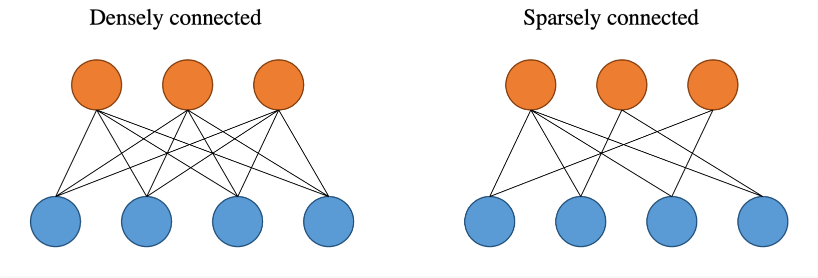

Figure: Dense vs Sparse networks[5]

[5] A. Alavi, *Using a VNN architecture*, https://amiralavi.net/blog/2018/07/29/vnn-implementation, [Online; accessed 26/02/2020], 2018.

How to learn connections?

Observed networks in data source

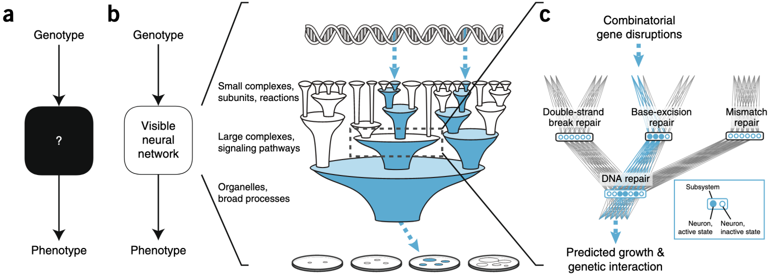

Figure: Protein interaction networks[6]

[6] J. Ma, M. K. Yu, S. Fong, *et al.*, "Using deep learning to model the hierarchical structure and function of a cell," *Nature methods*, vol. 15, no. 4, p. 290, 2018.

How to learn connections?

Pruning[7]

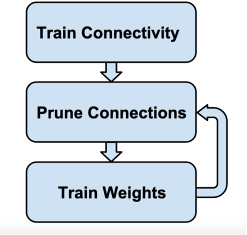

Figure: The three step pipeline

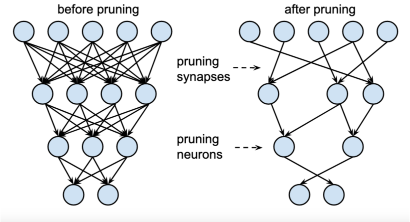

Figure: Pruning: before and after

[7] S. Han, J. Pool, J. Tran, *et al.*, "Learning both weights and connections for efficient neural network," in *Advances in neural information processing systems*, 2015, pp. 1135–1143.

How to learn connections?

Adaptive sparse connectivity

Algorithm 1: SET pseudocode

%Initialization;
initialize ANN model;
set ε and ζ;
for *each bipartite fully-connected (FC) layer of the ANN* **do**
    replace FC with a Sparse Connected (SC) layer having a Erdős-Rényi topology given by ε and Eq.1;
end
initialize training algorithm parameters;
%Training;
for *each training epoch e* **do**
    perform standard training procedure;
    perform weights update;
    for *each bipartite SC layer of the ANN* **do**
        remove a fraction ζ of the smallest positive weights;
        remove a fraction ζ of the largest negative weights;
        if *e is not the last training epoch* **then**
            add randomly new weights (connections) in the same amount as the ones removed previously;
        end
    end
end

Figure: Sparse Evolutionary Training (SET) algorithm[8]

[8] D. C. Mocanu, E. Mocanu, P. Stone, *et al.*, "Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science," *Nature Communications*, vol. 9, 2018. [Online]. Available: http://arxiv.org/abs/1707.04780 (visited on 02/26/2020).

SET: Encoding domain-specific information

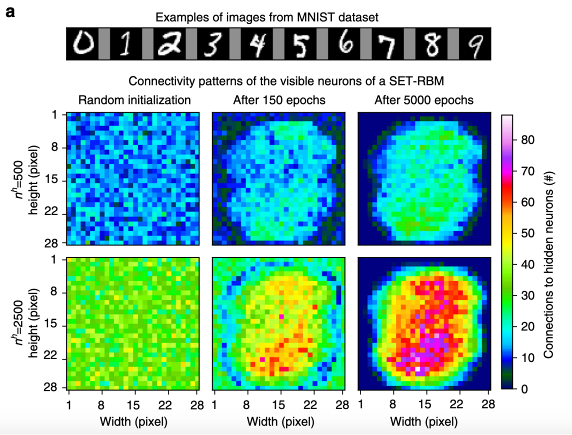

Figure: Input connections of an SET trained on MNIST[9]

[9] D. C. Mocanu, E. Mocanu, P. Stone, *et al.*, "Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science," *Nature Communications*, vol. 9, 2018. [Online]. Available: http://arxiv.org/abs/1707.04780 (visited on 02/26/2020).

Advantages

- Robustness[10]
- Better fit to lifelong learning[11]
- Task specific sub-networks[12]
- Help discover relationships in unstructured data[13]

[10] S. Ahmad and J. Hawkins, "How do neurons operate on sparse distributed representations? a mathematical theory of sparsity, neurons and active dendrites," *arXiv preprint arXiv:1601.00720*, 2016.

[11] Y. Cui, S. Ahmad, and J. Hawkins, "Continuous online sequence learning with an unsupervised neural network model," *Neural computation*, vol. 28, no. 11, pp. 2474–2504, 2016.

[12] S. Golkar, M. Kagan, and K. Cho, "Continual learning via neural pruning," *arXiv preprint arXiv:1903.04476*, 2019.

[13] D. C. Mocanu, E. Mocanu, P. Stone, *et al.*, "Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science," *Nature Communications*, vol. 9, 2018. [Online]. Available: `http://arxiv.org/abs/1707.04780` (visited on 02/26/2020).

Research directions

- Outperform dense models, but start by training dense models
- Sparse matrix multiplications limited in performance[14]

[14] S. Changpinyo, M. Sandler, and A. Zhmoginov, "The power of sparsity in convolutional neural networks," *arXiv preprint arXiv:1702.06257*, 2017.

References

1. D. C. Mocanu, E. Mocanu, P. Stone, *et al.*, "Scalable Training of Artificial Neural Networks with Adaptive Sparse Connectivity inspired by Network Science," *Nature Communications*, vol. 9, 2018. [Online]. Available: `http://arxiv.org/abs/1707.04780` (visited on 02/26/2020)
2. L. Souza, *A case for sparsity in neural networks: Pruning*, `https://numenta.com/blog/2019/08/30/case-for-sparsity-in-neural-networks-part-1-pruning`, [Online; accessed 26/02/2020], 2019
3. M. Klear, *The Sparse Future of Deep Learning*, en, Library Catalog: towardsdatascience.com, Dec. 2018. [Online]. Available: `https://towardsdatascience.com/the-sparse-future-of-deep-learning-bce05e8e094a` (visited on 02/26/2020)