MSc in Computational Cognitive Neuroscience

Goldsmiths
UNIVERSITY OF LONDON

Why the brain cannot be just many deep neural networks - A proof-of-concept study

Arno Vanegdom, Dr. Max Garagnani, Dr. Nikolay Nikolaev

Introduction

A fundamental cognitive function of the human brain is its ability to simultaneously hold in mind several items/symbols it had previously learned separately, by co-activating those items' internal neuronal representations. What we define as **superposition** (Figure 1).

A few studies have previously addressed its potential implementation in artificial neural networks in the context of the "superposition catastrophe" [1,2], the problem NNs have to overcome when trying to implement superposition. Those studies have provided conflicting results regarding the ability of NNs to Implement this function. Moreover, the NNs were tasked to superpose items ***during*** training.

This research project is separated in 2 parts. **In the 1st part**, we assessed the ability of standard artificial neural networks to implement the superposition of 2 items it had previously learned separately. **In the 2nd part**, we show how a neurobiologically constrained model of the brain [3], implementing a realistic synaptic plasticity learning rule, leads to the emergence of learned cell assemblies that can be subsequently co-activated and therefore allows superposition.

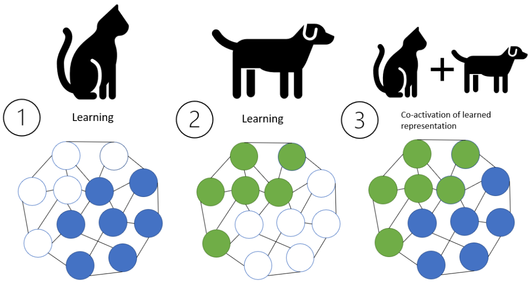


Figure 1, conceptual diagram of superposition

Objectives

- Assessing the ability (or lack of) of artificial intelligent systems to implement the fundamental cognitive function of superposition.
- Understanding the reasons, tied to the underlying inherent functioning and nature of such systems, determining and constraining this potential implementation.
- Understanding the mechanisms allowing the brain to implement superposition by proposing a biologically constrained computational model able to do so

Part 1 : Assessing the ability of ANNs to implement superposition

Hypothesis: ANNs cannot implement such function because of their distributed representations of items determined by their learning rule : **backpropagation**

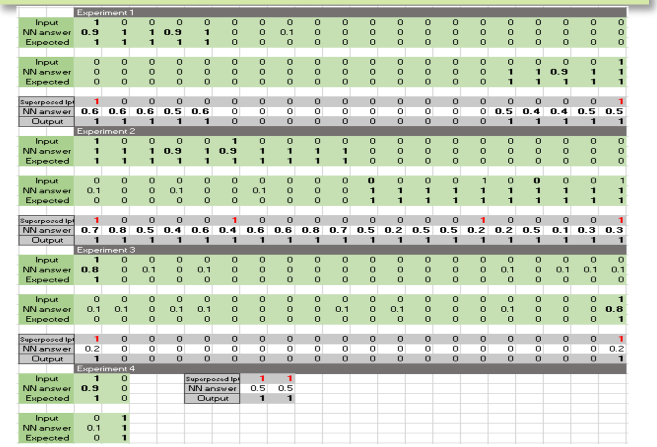
Results : The results confirmed our hypothesis; the NN was unable to correctly produce the superposed output in none of the trials across the four experiments (figure 2 and 3).

Part 1 : Assessing the ability of ANNs to implement superposition

Methods : We trained a one hidden-layer NN to associate 2 sets of 5 inputs patterns to an output pattern representing one of 2 category (or item).

After training, we presented the network a set of "superposed patterns" consisting of merged patterns of an input belong to one category superposed with an input pattern belonging to the other category. The NN ability to implement superposition was assessed whether it successfully produce the output pattern consisting of the superposition of both categories pattern (Figure 2).

We performed 4 experiments, each presenting variations in Input/output patterns characteristics. The 4th experiment consisted of incrementally decreasing the network's size in order to analyze the network computations, providing the reason why the NN was able or unable to implement superposition.



	Experiment 1																			
Input	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NN answer	0.9	1	1	0.9	1	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
Expected	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Input	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NN answer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0.9	1	1
Expected	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
Superposed Ipt	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NN answer	0.6	0.6	0.6	0.5	0.6	0	0	0	0	0	0	0	0	0	0	0.5	0.4	0.4	0.5	0.5
Output	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
	Experiment 2																			
Input	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NN answer	1	1	1	0.9	1	0.9	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Expected	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Input	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
NN answer	0.1	0	0	0.1	0	0	0.1	0	0	0	1	1	1	1	1	1	1	1	1	1
Expected	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Superposed Ipt	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1
NN answer	0.7	0.8	0.5	0.4	0.6	0.4	0.6	0.6	0.8	0.7	0.5	0.2	0.5	0.5	0.2	0.2	0.5	0.1	0.3	0.3
Output	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Experiment 3																			
Input	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NN answer	0.8	0	0.1	0	0.1	0	0	0	0	0	0	0	0	0	0	0.1	0	0.1	0.1	0.1
Expected	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Input	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NN answer	0.1	0.1	0	0.1	0.1	0	0	0	0	0.1	0	0.1	0	0	0	0.1	0	0	0	0.8
Expected	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Superposed Ipt	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NN answer	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2
Output	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	Experiment 4																			
Input	1	0			Superposed Ipt		1	1												
NN answer	0.9	0			NN answer		0.5	0.5												
Expected	1	0			Output		1	1												
Input	0	1																		
NN answer	0.1	1																		
Expected	0	1																		

Figure 2, examples of training patterns and superposition results

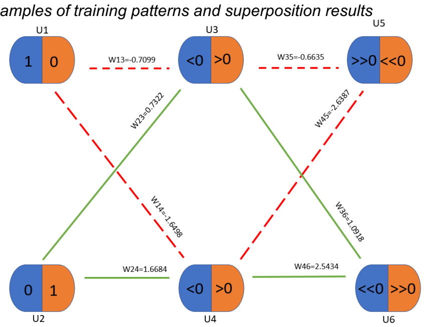


Figure 3, Experiment 4's learned representation and weighting scheme of the 6-units neural network

Part 2 : Bio-realistic computational model implementing superposition

Hypothesis : The model allows and implement superposition

Method : We will "train" a bio-realistic neural network model to learn internal representation of input patterns (items) according to a , "unsupervised", synaptic plasticity learning rule known to happen in the brain. Those internal representation (cell assemblies), can be co-activated and allow the model to implement superposition.

1. Bowers, J. S., Vankov, I. I., Damian, M. F., & Davis, C. J. (2014). Neural networks learn highly selective representations in order to overcome the superposition catastrophe. Psychological Review, 121(2), 248-261

2. Martin, N. D. (2021). Selectivity in neural networks (Doctoral dissertation, University of Bristol).

3. Garagnani, M., Wennekers, T., & Pulvermüller, F. (2009). Recruitment and Consolidation of Cell Assemblies for Words by Way of Hebbian Learning and Competition in a Multi-Layer Neural Network. Cognitive Computation, 1(2), 160–176.