

DECOMPOSITION ANALYSIS ON THE NEURAL NETWORK DESIGNS

Bui Gia Khanh *

Department of Physics Hanoi University of Science, Vietnam National University

Hanoi, Vietnam

{fujimiyaamane}@outlook.com

ABSTRACT

We revise and reevaluate the theory of artificial neural networks, with the cornerstone notion of *forward process* and *perceptron operations*. This is the framework which gives rise to *neural network architecture* and furthermore, the *deep learning architecture* of multilayer perceptrons (MLP). In particular, we construct a novel structures and analytical component sets, which serves to decompose the neural network architecture to the smallest working component, and within general principle of working.

1 INTRODUCTION

With the development of neurological models since the early 1940s [McCulloch, Pitts, 1943], the works done in 1950s to the ends of 1970s [Hebb, 1949], Rosenblatt (1958) [Bernard, Hoff, 1959] [Ivakhnenko, Lapa, 1965] [Shun'ichi Amari, 1967,1972], [Seppo, 1970] [Kely, 1960] by many people before the period called the AI winter, and progress made of which formed the discipline of *deep learning* up to present, machine learning has been mostly considered, and conducted, within the *neural network architecture*. This architecture is based on its most basic component of an *artificial neuron*, most often described and constructed as a *perceptron*. Many progresses has been made using this architecture involves *forward neural network* (FNN), *backpropagation principle*, *perceptron learning* and *multilayer perceptron* (MLP), *Hebbian learning*, *Hopfield networks*, *recurrent neural network* (RNN) and *long-short term memory* (LSTM) network, *convolutional neural network* (CNN), up to *encoder-decoder*, *belief networks*, and then *transformer* (*sub*)architecture, most prominent in applications of the state-of-the-art operations in the present time.

However, we still do not know entirely how the neural network architecture actually works, even from those successes conceded. The nature of the neural network architecture is partially unknown, typically described as a black box. Its working mechanism is well-known, but the decision making and information processing is untrackable. Works has been conducted to deal with this problem, specifically through viewing the neural network, and its subsequent construction called *deep learning* via theory and traceback to neurological concepts [Samuel et al., 2023] [Daniel, Sho, Boris, 2021v2] [Timothy, Konrad, 2019] [Zhanghao, Ding, 2021], or by practical experiments from different perspective [Jason et al., 2015] [Nguyen et al., 2019]¹. However, those researches conceived gained various insight, not an overall rework of the hypothesis, and most of the time is constrained by the rigid construction already presented. Theories are taken in turns, but some of the works focus on rather the general case, the bounds and conceptual limits, but not the exact phenomenological architecture itself. One example of such is the PAC-learning theory, PAC-Bayesian, VC-Theory, and else. An analysis of the actual optimizer (taken in terms of PAC(-Bayesian) theory *objective*) is very much the main objective of this paper.

Hence, this paper focus, and **goal**, is on the *deconstruction* and a novel construction of the main schema of artificial intelligence researches - the neural network architecture, and presents it in a more conceptual-practical and transparent (if able) cases. This also includes various insights, propositions, conjectures and incorporation of different concepts into the analysis. The obvious goal, *learning*, however, is not in the schedule to be made, or rather, is not considered in such case.

*Undergraduate student

¹We do not specify all papers related to such issues, but those are some highlights.

2 CONSTRUCTIONS

Neural network architecture was based around the construction of the singular processing unit, called *neuron*. This stems from neuroscience researches and knowledge, of which identifies the main processing construction is the *biological neuron* embedded in the development of the brain. McCulloch and Pitts first introduced, the first mathematical model of a biological neuron. This model consists of the classical separation with preactivation and *activation function*, specifically the Thresholding Logic Unit (TLU):

$$y_{in} = \sum_{i=1}^n y_{in_i} \quad \forall x_n \in \{0, 1\} \quad (1)$$

$$f(y_{in}) = \begin{cases} 1 & y_{in} \geq \theta \\ 0 & y_{in} < \theta \end{cases} \quad (2)$$

where θ is the threshold and y_{in} is the total net input signal received. The construction of this network follows propositional logics, and hence, it is theorized to simulate and construct *boolean logic handler*, for OR, AND, NAND or others. And with the parameter θ , its name as the TLU is common in electronics and computer systems. In principle, in their original paper *A logical calculus of the ideas immanent in nervous activity, 1943*, it is remarked that, their idea is to "[He] therefore attempted to record the behaviour of complicated nets in the notation of symbolic logic of propositions", in which the all-or-none principle of such time infers the absolute certainty of propositional logic. The goal is then apparent - to use symbolic and propositional logic to express the working of every network, and hence applies it to the main structure.

This idea of a computational unit with receiver and activation patterns continues from then. The later idea includes Rosenblatt's perceptron (1957,1958), of which was based on works of McCulloch, the MCP or TLU, Culbertson (1956), Minsky (1956), Hebb (1949), and some of Von Neumann (1951, 1956). In their construction, per example for a photo-perceptron (optical signals), the organization of a perceptron is pretty much complex, based on three components:

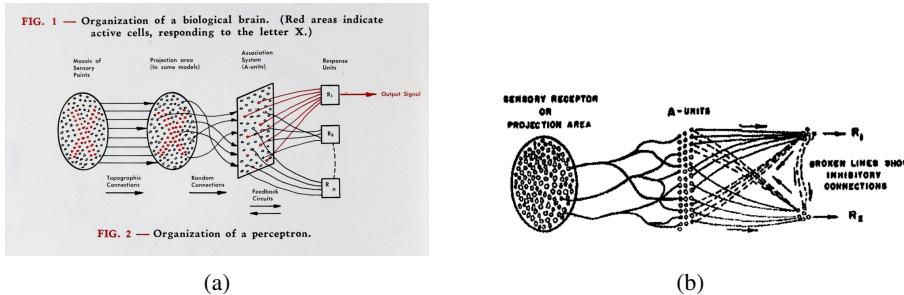


Figure 1: (a) Figure of the original organization of the biological model of the brain functions. (b) Specifically, note that it is specifically for optical case, but can be extended to others type. Furthermore, the layer between last A -unit and the response units, there exists a pattern of feedback loop.

1. The retina with localized connections, containing sensor points, called S -points.
2. The *projection area* A_I of which is described to be connected subsequently to the *sensor area* (this can be omitted, however, in case the next layer is directly connected to the retina), which receives a number of connections from the sensory points, and those formed a set called the *origin points* of such A -unit.
3. Between the projection area and the next layer here, we call as *association area*, connections are assumed to be random, scattered at random throughout the projection area.
4. The responses R_1, R_2, \dots are cells which respond in much the same fashion as the A -unit. Each response has a certain count of origin points located at random (similar to between A_I and A_{II}) in the A_{II} net.

Descriptively, we describe Rosenblatt's perceptron as followed.

Definition 2.1 (Rosenblatt, 1957). A perceptron (or linear classifier) is a function

$$\begin{aligned} \text{lin} : \mathbb{R}^d &\rightarrow \mathbb{R} \\ \mathbf{x} &\rightarrow \mathbf{w} \cdot \mathbf{x} + b \end{aligned} \tag{3}$$

where the parameters to be learned are $\mathbf{w} \in \mathbb{R}^d$ and $b \in \mathbb{R}$.

A intricate detail to observe, is the fact that between A_{II} and the receptive field, there exists feedback loops, directly connected back to each other, based on firing principles and thereof. There exists inhibitory and excitatory feedback connections, which separated the propagation phase into the *predominant* and *postdominant* phases. Effectively, this brings the idea of tangent optimization to it for one of the first network to do so. A note on Rosenblatt's perceptron, though, is that there are no *activation function* in the network. Most notably, Rosenblatt's and McCulloch's perceptrons are often deferred to as either linear classifier, or *binary classifier* on linear space [Freund, Schapire, 1999].

Albeit the role of activation function is dubious, it was presented in the very first TLU as a threshold unit, effectively transformed TLU into a 2-partition boundary unit within the parameter space. In one way or another, we can think of activation function belongs to a collection of main operating mechanism of which the neuron is thought to be modelled as.

2.1 MULTILAYER PERCEPTRON (MLP)

Moving on, the limits of perceptron is well-known, at least in its concurrent form. By the standard linear-binary classifier, the "XOR problem" is unsolvable by the typical perceptron system. In fact, any data of which is *linearly inseparable* will render perceptron useless, albeit we still have the perceptron cycling theorem [Block, Levin, 1970]. Thereby, a different architecture is involved, namely the *multilayer perceptron architecture* (MLP).

It is here that we should note that the name multilayer perceptron brings a lot of controversy to the name on itself. The name directly (and incorrectly) infers to the fact that it contains only one more component: a lot of perceptrons in layers. However, the name is talking about the architecture of a single neuron, and we can see the idea of **layers** having introduced earlier on in Rosenblatt's perceptron model. Rather, it is the fact that multilayer perceptron is the architecture that comprises many components into layers, but not every component is a perceptron, and rather, they are in one way or another, *neuron-like* processing unit. Many of today's architecture relies on such interpretation, including the latest structures. Those, for example, Transformer, CNN, and else, can be considered partially heuristic when considered to generality of the neurological processing operations, taking advantage of certain data configurations.

There are no definite definition of a multilayer perceptron. This is mainly because of its configuration and archetypical nature of being the 'way of construction' of meaningful compositions for smaller neural-like units, the name is given toward being an architecture instead on its own, in typical literature. [Ramchoun et al., 2016] [Roberts, Yaida, Hanin, 2021v2]

2.1.1 REMARK

Section 2 introduced the basic ground works for generally, all the currently constructed neural network system, per history and timeline of ideas. It leads to several questions:

- Activation function is essentially to the formation of MLP. However, their appearance is not generally dated. Before that, MCP and Rosenblatt's perceptron also has activation functions, albeit they are threshold units containing logical syllogism of either 0 or 1, except for ADALINE with linear activation function.

Multilayer perceptron is imperatively different, because it uses *non-linear* activation function. As we recalled, TLU can only solve the binary boundary problem, and while perceptron is better, it is ill-suited for problems that need non-linear solutions. Generally, it is true that linear unit can only interpret linear patterns and situations. Thereby, the non-linear activation function is considered a breakthrough, of which is prominent in the multilayer perceptron system. However, the question remains. What is the potential of an activation function?

- Early in section 2, we refer to components of multilayer perceptron as neural-like components. But what exactly are them? In a loose sense, they can be considered to be processing unit, with input-output flow of operation, and certain operation mutating the input to certain form - or rather, mapping it from the input space \mathcal{I} to the processor's space \mathcal{P} . But what is the true construction of such neural-like components. And further, what can be their concrete definition and theories?

3. What can be the representative mathematical object and operations that partially represents and interpret a single perceptron unit in actual system?
4. What is the *axiom* or assumption of the multilayer perceptron (for example, this question somewhat refers to Hebb's rule as one of the axiomatic condition of evolution of neural net, which then could be called as Hebbian neural architecture)?
5. What is exactly meant by *neural-like components*? From Rosenblatt to Marvin, their idea of components networks or *Society of Mind* in Marvin's case, but what exactly constitute the neural-like properties, and what constitute those components to works together?

3 A STRUCTURAL ANALYSIS

REFERENCES

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, 1958. URL <https://api.semanticscholar.org/CorpusID:12781225>.

A APPENDIX

You may include other additional sections here.