Title: A Logical Calculus of the Ideas Immanent in Nervous Activity

URL: https://en.wikipedia.org/wiki/A_Logical_Calculus_of_the_Ideas_Immanent_in_Nervous_Activity

PageID: 78113392

Categories: Category:Artificial neural networks, Category:Computer science papers, Category:History of artificial intelligence, Category:Machine learning

---

"A Logical Calculus of the Ideas Immanent in Nervous Activity" is a 1943 article written by Warren McCulloch and Walter Pitts . The paper, published in the journal The Bulletin of Mathematical Biophysics , proposed a mathematical model of the nervous system as a network of simple logical elements, later known as artificial neurons, or McCulloch-Pitts neurons . These neurons receive inputs, perform a weighted sum, and fire an output signal based on a threshold function. By connecting these units in various configurations, McCulloch and Pitts demonstrated that their model could perform all logical functions.

It is a seminal work in cognitive science , computational neuroscience , computer science , and artificial intelligence . It was a foundational result in automata theory . John von Neumann cited it as a significant result.

Mathematics

The artificial neuron used in the original paper is slightly different from the modern version. They considered neural networks that operate in discrete steps of time t = 0 , 1 , … {\displaystyle t=0,1,\dots } .

The neural network contains a number of neurons. Let the state of a neuron i {\displaystyle i} at time t {\displaystyle t} be N i (t) {\displaystyle N_{i}(t)} . The state of a neuron can either be 0 or 1, standing for "not firing" and "firing". Each neuron also has a firing threshold θ {\displaystyle \theta } , such that it fires if the total input exceeds the threshold.

Each neuron can connect to any other neuron (including itself) with positive synapses (excitatory) or negative synapses (inhibitory). That is, each neuron can connect to another neuron with a weight w {\displaystyle w} taking an integer value. A peripheral afferent is a neuron with no incoming synapses.

We can regard each neural network as a directed graph , with the nodes being the neurons, and the directed edges being the synapses. A neural network has a circle or a circuit if there exists a directed circle in the graph.

Let w i j (t) {\displaystyle w_{ij}(t)} be the connection weight from neuron j {\displaystyle j} to neuron i {\displaystyle i} at time t {\displaystyle t} , then its next state is N i (t + 1) = H (∑ j = 1 n w i j (t) N j (t) − θ i (t)) , {\displaystyle N_{i}(t+1)=H\left(\sum _{j=1}^{n}w_{ij}(t)N_{j}(t)-\theta _{i}(t)\right),} where H {\displaystyle H} is the Heaviside step function (outputting 1 if the input is greater than or equal to 0, and 0 otherwise).

Symbolic logic

The paper used, as a logical language for describing neural networks, "Language II" from The Logical Syntax of Language by Rudolf Carnap with some notations taken from Principia Mathematica by Alfred North Whitehead and Bertrand Russell . Language II covers substantial parts of classical mathematics, including real analysis and portions of set theory.

To describe a neural network with peripheral afferents N 1 , N 2 , … , N p {\displaystyle N_{1},N_{2},\dots ,N_{p}} and non-peripheral afferents N p + 1 , N p + 2 , … , N n {\displaystyle N_{p+1},N_{p+2},\dots ,N_{n}} they considered logical predicate of form P r (N 1 , N 2 , … , N p , t) {\displaystyle Pr(N_{1},N_{2},\dots ,N_{p},t)} where P r {\displaystyle Pr} is a first-order logic predicate function (a function that outputs a boolean), N 1 , … , N p {\displaystyle N_{1},\dots

,N_{p}} are predicates that take t {\displaystyle t} as an argument, and t {\displaystyle t} is the only free variable in the predicate. Intuitively speaking, N 1 , … , N p {\displaystyle N_{1},\dots ,N_{p}} specifies the binary input patterns going into the neural network over all time, and P r (N 1 , N 2 , … , N n , t) {\displaystyle Pr(N_{1},N_{2},\dots ,N_{n},t)} is a function that takes some binary input patterns, and constructs an output binary pattern P r (N 1 , N 2 , … , N n , 0) , P r (N 1 , N 2 , … , N n , 1) , … {\displaystyle Pr(N_{1},N_{2},\dots ,N_{n},0),Pr(N_{1},N_{2},\dots ,N_{n},1),\dots } .

A logical sentence P r (N 1 , N 2 , … , N n , t) {\displaystyle Pr(N_{1},N_{2},\dots ,N_{n},t)} is realized by a neural network iff there exists a time-delay T ≥ 0 {\displaystyle T\geq 0} , a neuron i {\displaystyle i} in the network, and an initial state for the non-peripheral neurons N p + 1 (0) , … , N n (0) {\displaystyle N_{p+1}(0),\dots ,N_{n}(0)} , such that for any time t {\displaystyle t} , the truth-value of the logical sentence is equal to the state of the neuron i {\displaystyle i} at time t + T {\displaystyle t+T} . That is, ∀ t = 0 , 1 , 2 , … , P r (N 1 , N 2 , … , N p , t) = N i (t + T) {\displaystyle \forall t=0,1,2,\dots ,\quad Pr(N_{1},N_{2},\dots ,N_{p},t)=N_{i}(t+T)}

Equivalence

In the paper, they considered some alternative definitions of artificial neural networks, and have shown them to be equivalent, that is, neural networks under one definition realizes precisely the same logical sentences as neural networks under another definition.

They considered three forms of inhibition: relative inhibition, absolute inhibition, and extinction. The definition above is relative inhibition. By "absolute inhibition" they meant that if any negative synapse fires, then the neuron will not fire. By "extinction" they meant that if at time t {\displaystyle t} , any inhibitory synapse fires on a neuron i {\displaystyle i} , then θ i (t + j) = θ i (0) + b j {\displaystyle \theta _{i}(t+j)=\theta _{i}(0)+b_{j}} for j = 1 , 2 , 3 , … {\displaystyle j=1,2,3,\dots } , until the next time an inhibitory synapse fires on i {\displaystyle i} . It is required that b j = 0 {\displaystyle b_{j}=0} for all large j {\displaystyle j} .

Theorem 4 and 5 state that these are equivalent.

They considered three forms of excitation: spatial summation, temporal summation, and facilitation. The definition above is spatial summation (which they pictured as having multiple synapses placed close together, so that the effect of their firing sums up). By "temporal summation" they meant that the total incoming signal is ∑ τ = 0 T ∑ j = 1 n w i j (t) N j (t − τ) {\displaystyle \sum _{\tau =0}^{T}\sum _{j=1}^{n}w_{ij}(t)N_{j}(t-\tau)} for some T ≥ 1 {\displaystyle T\geq 1} . By "facilitation" they meant the same as extinction, except that b j ≤ 0 {\displaystyle b_{j}\leq 0} . Theorem 6 states that these are equivalent.

They considered neural networks that do not change, and those that change by Hebbian learning . That is, they assume that at t = 0 {\displaystyle t=0} , some excitatory synaptic connections are not active. If at any t {\displaystyle t} , both N i (t) = 1 , N j (t) = 1 {\displaystyle N_{i}(t)=1,N_{j}(t)=1} , then any latent excitatory synapse between i , j {\displaystyle i,j} becomes active. Theorem 7 states that these are equivalent.

Logical expressivity

They considered "temporal propositional expressions" (TPE), which are propositional formulas with one free variable t {\displaystyle t} . For example, N 1 (t) ∨ N 2 (t) ∧ ¬ N 3 (t) {\displaystyle N_{1}(t)\vee N_{2}(t)\wedge \neg N_{3}(t)} is such an expression. Theorem 1 and 2 together showed that neural nets without circles are equivalent to TPE.

For neural nets with loops, they noted that "realizable P r {\displaystyle Pr} may involve reference to past events of an indefinite degree of remoteness". These then encodes for sentences like "There was some x such that x was a ψ" or (∃ x) (ψ x) {\displaystyle (\exists x)(\psi x)} . Theorems 8 to 10 showed that neural nets with loops can encode all first-order logic with equality and conversely, any looped neural networks is equivalent to a sentence in first-order logic with equality, thus showing that they are equivalent in logical expressiveness.

As a remark, they noted that a neural network, if furnished with a tape, scanners, and write-heads, is equivalent to a Turing machine , and conversely, every Turing machine is equivalent to some such neural network. Thus, these neural networks are equivalent to Turing computability , Church 's

lambda-definability , and Kleene 's primitive recursiveness .

Context

Previous work

The paper built upon several previous strands of work.

In the symbolic logic side, it built on the previous work by Carnap, Whitehead, and Russell. This was contributed by Walter Pitts, who had a strong proficiency with symbolic logic. Pitts provided mathematical and logical rigor to McCulloch’s vague ideas on psychons (atoms of psychological events) and circular causality.

In the neuroscience side, it built on previous work by the mathematical biology research group centered around Nicolas Rashevsky , of which McCulloch was a member. The paper was published in the Bulletin of Mathematical Biophysics , which was founded by Rashevsky in 1939. During the late 1930s, Rashevsky's research group was producing papers that had difficulty publishing in other journals at the time, so Rashevsky decided to found a new journal exclusively devoted to mathematical biophysics.

Also in the Rashevsky's group was Alston Scott Householder , who in 1941 published an abstract model of the steady-state activity of biological neural networks. The model, in modern language, is an artificial neural network with ReLU activation function. In a series of papers, Householder calculated the stable states of very simple networks: a chain, a circle, and a bouquet . Walter Pitts' first two papers formulated a mathematical theory of learning and conditioning. The next three were mathematical developments of Householder’s model.

In 1938, at age 15, Pitts ran away from home in Detroit and arrived in the University of Chicago . Later, he walked into Rudolf Carnap's office with Carnap's book filled with corrections and suggested improvements. He started studying under Carnap and attending classes during 1938--1943. He wrote several early papers on neuronal network modelling and regularly attended Rashevsky's seminars in theoretical biology. The seminar attendants included Gerhard von Bonin and Householder. In 1940, von Bonin introduced Lettvin to McCulloch. In 1942, both Lettvin and Pitts had moved in with McCulloch's home.

McCulloch had been interested in circular causality from studies with causalgia after amputation, epileptic activity of surgically isolated brain, and Lorente de Nò 's research showing recurrent neural networks are needed to explain vestibular nystagmus . He had difficulty with treating circular causality until Pitts demonstrated how it can be treated by the appropriate mathematical tools of modular arithmetics and symbolic logic.

Both authors' affiliation in the article was given as "University of Illinois, College of Medicine, Department of Psychiatry at the Illinois Neuropsychiatric Institute, University of Chicago, Chicago, U.S.A."

Subsequent work

It was a foundational result in automata theory . John von Neumann cited it as a significant result. This work led to work on neural networks and their link to finite automata . Kleene introduced the term "regular" for "regular language" in a 1951 technical report, where Kleene proved that regular languages are all that could be generated by neural networks, among other results. The term "regular" was meant to be suggestive of "regularly occurring events" that the neural net automaton must process and respond to.

Marvin Minsky was influenced by McCulloch, built an early example of neural network SNARC (1951), and did a PhD thesis on neural networks (1954).

McCulloch was the chair to the ten Macy conferences (1946--1953) on "Circular Causal and Feedback Mechanisms in Biological and Social Systems". This was a key event in the beginning of cybernetics , and what later became known as cognitive science . Pitts also attended the conferences.

In the 1943 paper, they described how memories can be formed by a neural network with loops in it, or alterable synapses, which are operating over time, and implements logical universals -- "there

exists" and "for all". This was generalized for spatial objects, such as geometric figures, in their 1947 paper How we know universals . Norbert Wiener found this a significant evidence for a general method for how animals recognizing objects, by scanning a scene from multiple transformations and finding a canonical representation. He hypothesized that this "scanning" activity is clocked by the alpha wave , which he mistakenly thought was tightly regulated at 10 Hz (instead of the 8 -- 13 Hz as modern research shows).

McCulloch worked with Manuel Blum in studying how a neural network can be "logically stable", that is, can implement a boolean function even if the activation thresholds of individual neurons are varied. They were inspired by the problem of how the brain can perform the same functions, such as breathing, under influence of caffeine or alcohol , which shifts the activation threshold over the entire brain.

See also

Artificial neural network

Perceptron

Connectionism

Principia Mathematica

History of artificial neural networks

References