

Connectionism

From Wikipedia, the free encyclopedia

Connectionism is a set of approaches in the fields of artificial intelligence, cognitive psychology, cognitive science, neuroscience, and philosophy of mind, that models mental or behavioral phenomena as the emergent processes of *interconnected networks of simple units*. The term was introduced by Donald Hebb in 1940s.^[1] There are many forms of connectionism, but the most common forms use neural network models.

Basic principles

The central connectionist principle is that mental phenomena can be described by interconnected networks of simple and often uniform units. The form of the connections and the units can vary from model to model. For example, units in the network could represent neurons and the connections could represent synapses.

Spreading activation

In most connectionist models, networks change over time. A closely related and very common aspect of connectionist models is *activation*. At any time, a unit in the network has an activation, which is a numerical value intended to represent some aspect of the unit. For example, if the units in the model are neurons, the activation could represent the probability that the neuron would generate an action potential spike. Activation typically spreads to all the other units connected to it. Spreading activation is always a feature of neural network models, and it is very common in connectionist models used by cognitive psychologists.

Neural networks

Neural networks are by far the most commonly used connectionist model today. Though there are a large variety of neural network models, they almost always follow two basic principles regarding the mind:

1. Any mental state can be described as an (N)-dimensional vector of numeric activation values over neural units in a network.
2. Memory is created by modifying the strength of the connections between neural units. The connection strengths, or "weights", are generally represented as an $N \times N$ matrix.

Most of the variety among neural network models comes from:

- *Interpretation of units*: Units can be interpreted as neurons or groups of neurons.
- *Definition of activation*: Activation can be defined in a variety of ways. For example, in a Boltzmann machine, the activation is interpreted as the probability of generating an action potential spike, and is determined via a logistic function on the sum of the inputs to a unit.
- *Learning algorithm*: Different networks modify their connections differently. In general, any mathematically defined change in connection weights over time is referred to as the "learning algorithm".

Connectionists are in agreement that recurrent neural networks (directed networks wherein connections of the network can form a directed cycle) are a better model of the brain than feedforward neural networks (directed networks with no cycles, called DAG). Many recurrent connectionist models also incorporate dynamical systems theory. Many researchers, such as the connectionist Paul Smolensky, have argued that connectionist models will evolve toward fully continuous, high-dimensional, non-linear, dynamic systems approaches.

Biological realism

The neural network branch of connectionism suggests that the study of mental activity is really the study of neural systems. This links connectionism to neuroscience, and models involve varying degrees of biological realism. Connectionist work in general need not be biologically realistic, but some neural network researchers, computational neuroscientists, try to model the biological aspects of natural neural systems very closely in so-called "neuromorphic networks". Many authors find the clear link between neural activity and cognition to be an appealing aspect of connectionism.

Learning

The weights in a neural network are adjusted according to some learning rule or algorithm, such as Hebbian learning. Thus, connectionists have created many sophisticated learning procedures for neural networks. Learning always involves modifying the connection weights. In general, these involve mathematical formulas to determine the change in weights when given sets of data consisting of activation vectors for some subset of the neural units.

By formalizing learning in such a way, connectionists have many tools. A very common strategy in connectionist learning methods is to incorporate gradient descent over an error surface in a space defined by the weight matrix. All gradient descent learning in connectionist models involves changing each weight by the partial derivative of the error surface with respect to the weight. Backpropagation (BP), first made popular in the 1980s, is probably the most commonly known connectionist gradient descent algorithm today.

History

Connectionism can be traced to ideas more than a century old, which were little more than speculation until the mid-to-late 20th century. Through his work on the structure of the nervous system for which he won the Nobel Prize in 1906, the Spanish Santiago Ramón y Cajal established the basis for studies of neural networks, but it wasn't until the 1980s that connectionism became a popular perspective among scientists.

Parallel distributed processing

The prevailing connectionist approach today was originally known as **parallel distributed processing** (PDP). It was an artificial neural network approach that stressed the parallel nature of neural processing, and the distributed nature of neural representations. It provided a general mathematical framework for researchers to operate in. The framework involved eight major aspects:

- A set of *processing units*, represented by a set of integers.
- An *activation* for each unit, represented by a vector of time-dependent functions.
- An *output function* for each unit, represented by a vector of functions on the activations.
- A *pattern of connectivity* among units, represented by a matrix of real numbers indicating connection strength.
- A *propagation rule* spreading the activations via the connections, represented by a function on the output of the units.
- An *activation rule* for combining inputs to a unit to determine its new activation, represented by a function on the current activation and propagation.
- A *learning rule* for modifying connections based on experience, represented by a change in the weights based on any number of variables.
- An *environment* that provides the system with experience, represented by sets of activation vectors for some subset of the units.

A lot of the research that led to the development of PDP was done in the 1970s, but PDP became popular in the 1980s with the release of the books *Parallel Distributed Processing: Explorations in the Microstructure of Cognition - Volume 1 (foundations)* and *Volume 2 (Psychological and Biological Models)*, by James L. McClelland, David E. Rumelhart and the PDP Research Group. The books are now considered seminal connectionist works, and it is now common to fully equate PDP and connectionism, although the term "connectionism" is not used in the books.

Earlier work

PDP's direct roots were the perceptron theories of researchers such as Frank Rosenblatt from the 1950s and 1960s. But perceptron models were made very unpopular by the book *Perceptrons* by Marvin Minsky and Seymour Papert, published in 1969. It demonstrated the limits on the sorts of functions that single-layered (no hidden layer) perceptrons can calculate, showing that even simple functions like the exclusive disjunction (XOR) could not be handled properly. The PDP books overcame this limitation by showing that multi-level, non-linear

neural networks were far more robust and could be used for a vast array of functions.^[2]

Many earlier researchers advocated connectionist style models, for example in the 1940s and 1950s, Warren McCulloch and Walter Pitts (MP neuron), Donald Olding Hebb, and Karl Lashley. McCulloch and Pitts showed how neural systems could implement first-order logic: Their classic paper "A Logical Calculus of Ideas Immanent in Nervous Activity" (1943) is important in this development here. They were influenced by the important work of Nicolas Rashevsky in the 1930s. Hebb contributed greatly to speculations about neural functioning, and proposed a learning principle, Hebbian learning, that is still used today. Lashley argued for distributed representations as a result of his failure to find anything like a localized engram in years of lesion experiments.

Connectionism apart from PDP

Though PDP is the dominant form of connectionism, other theoretical work should also be classified as connectionist.

Many connectionist principles can be traced to early work in psychology, such as that of William James.^[3] Psychological theories based on knowledge about the human brain were fashionable in the late 19th century. As early as 1869, the neurologist John Hughlings Jackson argued for multi-level, distributed systems. Following from this lead, Herbert Spencer's *Principles of Psychology*, 3rd edition (1872), and Sigmund Freud's *Project for a Scientific Psychology* (composed 1895) propounded connectionist or proto-connectionist theories. These tended to be speculative theories. But by the early 20th century, Edward Thorndike was experimenting on learning that posited a connectionist type network.

In the 1950s, Friedrich Hayek proposed that spontaneous order in the brain arose out of decentralized networks of simple units. Hayek's work was rarely cited in the PDP literature until recently

Another form of connectionist model was the relational network framework developed by the linguist Sydney Lamb in the 1960s. Relational networks have been only used by linguists, and were never unified with the PDP approach. As a result, they are now used by very few researchers.

There are also hybrid connectionist models, mostly mixing symbolic representations with neural network models. The hybrid approach has been advocated by some researchers (such as Ron Sun).

Connectionism vs. computationalism debate

As connectionism became increasingly popular in the late 1980s, there was a reaction to it by some researchers, including Jerry Fodor, Steven Pinker and others. They argued that connectionism, as it was being developed, was in danger of obliterating what they saw as the progress being made in the fields of cognitive science and psychology by the classical approach of computationalism. Computationalism is a specific form of cognitivism that argues that mental activity is computational, that is, that the mind operates by performing purely formal operations on symbols, like a Turing machine. Some researchers argued that the trend in connectionism was a reversion toward associationism and the abandonment of the idea of a language of thought, something they felt was mistaken. In contrast, it was those very tendencies that made connectionism attractive for other researchers.

Connectionism and computationalism need not be at odds, but the debate in the late 1980s and early 1990s led to opposition between the two approaches. Throughout the debate, some researchers have argued that connectionism and computationalism are fully compatible, though full consensus on this issue has not been reached. The differences between the two approaches that are usually cited are the following:

- Computationalists posit symbolic models that are structurally similar to underlying brain structure, whereas connectionists engage in "low-level" modeling, trying to ensure that their models resemble neurological structures.
- Computationalists in general focus on the structure of explicit symbols (mental models) and syntactical rules for their internal manipulation, whereas connectionists focus on learning from environmental stimuli and storing this information in a form of connections between neurons.
- Computationalists believe that internal mental activity consists of manipulation of explicit symbols, whereas connectionists

believe that the manipulation of explicit symbols is a poor model of mental activity.

- Computationalists often posit domain specific symbolic subsystems designed to support learning in specific areas of cognition (e.g., language, intentionality, number), whereas connectionists posit one or a small set of very general learning mechanisms.

But, despite these differences, some theorists have proposed that the connectionist architecture is simply the manner in which the symbol manipulation system happens to be implemented in the organic brain. This is logically possible, as it is well known that connectionist models can implement symbol manipulation systems of the kind used in computationalist models,^[citation needed] as indeed they must be able if they are to explain the human ability to perform symbol manipulation tasks. But the debate rests on whether this symbol manipulation forms the foundation of cognition in general, so this is not a potential vindication of computationalism. Nonetheless, computational descriptions may be helpful high-level descriptions of cognition of logic, for example.

The debate largely centered on logical arguments about whether connectionist networks were capable of producing the syntactic structure observed in this sort of reasoning. This was later achieved,^[citation needed] although using processes unlikely to be possible in the brain,^[citation needed] thus the debate persisted. Today, progress in neurophysiology, and general advances in the understanding of neural networks, has led to the successful modelling of a great many of these early problems, and the debate about fundamental cognition has, thus, largely been decided among neuroscientists in favor of connectionism.^[citation needed] However, these fairly recent developments have yet to reach consensus acceptance among those working in other fields, such as psychology or philosophy of mind.

Part of the appeal of computational descriptions is that they are relatively easy to interpret, and thus may be seen as contributing to our understanding of particular mental processes, whereas connectionist models are in general more opaque, to the extent that they may be describable only in very general terms (such as specifying the learning algorithm, the number of units, etc.), or in unhelpfully low-level terms. In this sense connectionist models may instantiate, and thereby provide evidence for, a broad theory of cognition (i.e., connectionism), without representing a helpful

theory of the particular process that is being modelled. In this sense the debate might be considered as to some extent reflecting a mere difference in the level of analysis in which particular theories are framed.

The recent popularity of dynamical systems in philosophy of mind have added a new perspective on the debate; some authors now argue that any split between connectionism and computationalism is more conclusively characterized as a split between computationalism and dynamical systems.

Notes

1. Elman, Jeffrey L.; et al. (1996). "Preface". *Rethinking Innateness: A Connectionist Perspective on Development (Neural Network Modeling and Connectionism)*. A Bradford Book. ISBN 978-0262550307. connectionism (a term introduced by Donald Hebb in 1940s, and the name we adopt here)
2. Hornik, K.; Stinchcombe, M.; White, H. (1989). "Multilayer feedforward networks are universal approximators". *Neural Networks* **2** (5): 359. doi:10.1016/0893-6080(89)90020-8.
3. Anderson, James A.; Rosenfeld, Edward (1989). "Chapter 1: (1890) William James Psychology (Brief Course)". *Neurocomputing: Foundations of Research*. A Bradford Book. p. 1. ISBN 978-0262510486.

References

- Rumelhart, D.E., J.L. McClelland and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, Cambridge, MA: MIT Press
- McClelland, J.L., D.E. Rumelhart and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models*, Cambridge, MA: MIT Press
- Pinker, Steven and Mehler, Jacques (1988). *Connections and Symbols*, Cambridge MA: MIT Press.

- Jeffrey L. Elman, Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, Kim Plunkett (1996). *Rethinking Innateness: A connectionist perspective on development*, Cambridge MA: MIT Press.
- Marcus, Gary F. (2001). *The Algebraic Mind: Integrating Connectionism and Cognitive Science (Learning, Development, and Conceptual Change)*, Cambridge, MA: MIT Press