As Machines Get Smarter, Evidence They Learn Like Us

By Natalie Wolchover

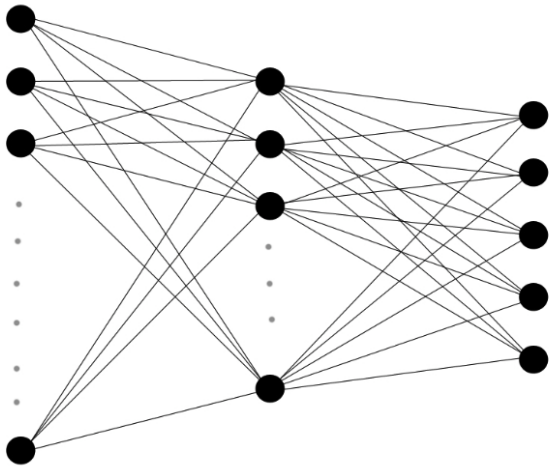
Studies suggest that computer models called neural networks, which are used in a growing number of applications, may 'learn' to recognize patterns in data using the same algorithms as the human brain.

The brain performs its canonical task — learning — by tweaking its myriad connections according to a secret set of rules. To unlock these secrets, scientists 30 years ago began developing computer models that try to replicate the learning process. Now, a growing number of experiments are revealing that these models behave strikingly similar to actual brains when performing certain tasks. Researchers say the similarities suggest a basic correspondence between the brains' and computers' underlying learning algorithms.

The algorithm used by a computer model called the Boltzmann machine, invented by Geoffrey Hinton and Terry Sejnowski in 1983, appears particularly promising as a simple theoretical explanation of a number of brain processes, including development, memory formation, object and sound recognition, and the sleep-wake cycle.

“It’s the best possibility we really have for understanding the brain at present,” said Sue Becker, a professor of psychology, neuroscience, and behavior at McMaster University in Hamilton, Ontario. “I don’t know of a model that explains a wider range of phenomena in terms of learning and the structure of the brain.”

Hinton, a pioneer in the field of artificial intelligence, has always wanted to understand the rules governing when the brain beefs a connection up and when it whittles one down — in short, the algorithm for how we learn. “It seemed to me if you want to understand something, you need to be able to build one,” he said. Following the reductionist approach of physics, his plan was to construct simple computer models of the brain that employed a variety of learning algorithms and “see which ones work,” said Hinton, who splits his time between the University of Toronto, where he is a professor of computer science, and Google.



Multilayer neural networks consist of layers of artificial neurons with weighted connections between them. Input data fed into the network sends a cascade of signals through the layers, and a learning algorithm dictates whether to increase or decrease the weight of each connection. The result is a network more attuned to the patterns that exist in data.

During the 1980s and 1990s, Hinton — the great-great-grandson of the 19th-century logician George Boole, whose work is the foundation of modern computer science — invented or co-invented a collection of machine learning algorithms. The algorithms, which tell computers how to learn from data, are used in computer models called artificial neural networks — webs of interconnected virtual neurons that transmit signals to their neighbors by switching on and off, or “firing.” When data are fed into the network, setting off a cascade of firing activity, the algorithm determines based on the firing patterns whether to increase or decrease the weight of the connection, or synapse, between each pair of neurons.

For decades, many of Hinton’s computer models languished. But thanks to advances in computing power, scientists’ understanding of the brain and the algorithms themselves, neural networks are playing an increasingly important role in neuroscience. Sejnowski, head of the Computational Neurobiology Laboratory at the Salk Institute for Biological Studies in La Jolla, Calif., said: “Thirty years ago, we had very crude ideas; now we are beginning to test some of those ideas.”

Brain Machines

Hinton’s early attempts at replicating the brain were limited. Computers could run his learning algorithms on small neural networks, but scaling the models up quickly overwhelmed the processors. In 2005, Hinton discovered that if he sectioned his neural networks into layers and ran

the algorithms on them one layer at a time, which approximates the brain's structure and development, the process became more efficient.

Although Hinton published his discovery in two top journals, neural networks had fallen out of favor by then, and "he was struggling to get people interested," said Li Deng, a principal researcher at Microsoft Research in Washington state. Deng, however, knew Hinton and decided to give his "deep learning" method a try in 2009, quickly seeing its potential. In the years since, the theoretical learning algorithms have been put to practical use in a surging number of applications, such as the Google Now personal assistant and the voice search feature on Microsoft Windows phones.

One of the most promising of these algorithms, the Boltzmann machine, bears the name of 19th century Austrian physicist Ludwig Boltzmann, who developed the branch of physics dealing with large numbers of particles, known as statistical mechanics. Boltzmann discovered an equation giving the probability of a gas of molecules having a particular energy when it reaches equilibrium. Replace molecules with neurons, and the Boltzmann machine, as it fires, converges on exactly the same equation.

The synapses in the network start out with a random distribution of weights, and the weights are gradually tweaked according to a remarkably simple procedure: The neural firing pattern generated while the machine is being fed data (such as images or sounds) is compared with random firing activity that occurs while the input is turned off.


Geoffrey Hinton, a pioneer in the field of artificial intelligence, thinks the best approach to understanding how brains learn is to try to build computers that learn in the same way. "You inevitably discover a lot about the computational issues, and you discover them at a level of understanding that psychologists don't have," he said.

Each virtual synapse tracks both sets of statistics. If the neurons it connects fire in close sequence more frequently when driven by data than when they are firing randomly, the weight of the synapse is increased by an amount proportional to the difference. But if two neurons more often fire together during random firing than data-driven firing, the synapse connecting them is too thick and consequently is weakened.

The most commonly used version of the Boltzmann machine works best when it is "trained," or fed thousands of examples of data, one layer at a time. First, the bottom layer of the network receives raw data representing pixelated images or multitonal sounds, and like retinal cells, neurons fire if they detect contrasts in their patch of the data, such as a switch from light to dark. Firing may

trigger connected neurons to fire, too, depending on the weight of the synapse between them. As the firing of pairs of virtual neurons is repeatedly compared with background firing statistics, meaningful relationships between neurons are gradually established and reinforced. The weights of the synapses are honed, and image or sound categories become ingrained in the connections. Each subsequent layer is trained the same way, using input data from the layer below.

If a picture of a car is fed into a neural network trained to detect specific objects in images, the lower layer fires if it detects a contrast, which would indicate an edge or endpoint. These neurons' signals travel to high level neurons, which detect corners, parts of wheels, and so on. In the top layer, there are neurons that fire only if the image contains a car.

"The magic thing that happens is it's able to generalize," said Yann LeCun, director of the Center for Data Science at New York University. "If you show it a car it has never seen before, if it has some common shape or aspect to all the cars you showed it during training, it can determine it's a car."

Neural networks have recently hit their stride thanks to Hinton's layer-by-layer training regimen, the use of high-speed computer chips called graphical processing units, and an explosive rise in the number of images and recorded speech available to be used for training. The networks can now correctly recognize about 88 percent of the words spoken in normal, human, English-language conversations, compared with about 96 percent for an average human listener. They can identify cars and thousands of other objects in images with similar accuracy and in the past three years have come to dominate machine learning competitions.

Build-a-Brain

No one knows how to directly ascertain the brain's learning rules, but there are many highly suggestive similarities between the brain's behavior and that of the Boltzmann machine.

Both learn with no supervision other than the patterns that naturally exist in data. "You don't get millions of examples of your mother telling you what's in an image," Hinton said. "You have to learn to recognize things without anybody telling you what the things are. Then after you learn the categories, people tell you the names of these categories. So kids learn about dogs and cats and then they learn that dogs are called 'dogs' and cats are called 'cats.' "

Adult brains are less malleable than juvenile ones, much as a Boltzmann machine trained with 100,000 car images won't change much upon seeing another: Its synapses already have the correct weights to categorize a car. And yet, learning never ends. New information can still be integrated into the structure of both brains and Boltzmann machines.

Over the past five to 10 years, studies of brain activity during sleep have provided some of the first direct evidence that the brain employs a Boltzmann-like learning algorithm in order to integrate new information and memories into its structure. Neuroscientists have long known that sleep plays an important role in memory consolidation, helping to integrate newly learned information. In 1995, Hinton and colleagues proposed that sleep serves the same function as the baseline component of the algorithm, the rate of neural activity in the absence of input.

"What you're doing during sleep is you're just figuring out the base rate," Hinton said. "You're figuring out how correlated would these neurons be if the system were running by itself. And then if the neurons are more correlated than that, increase the weight between them. And if they're less correlated than that, decrease the weight between them."

At the level of the synapses, "this algorithm can be implemented in several different ways," said

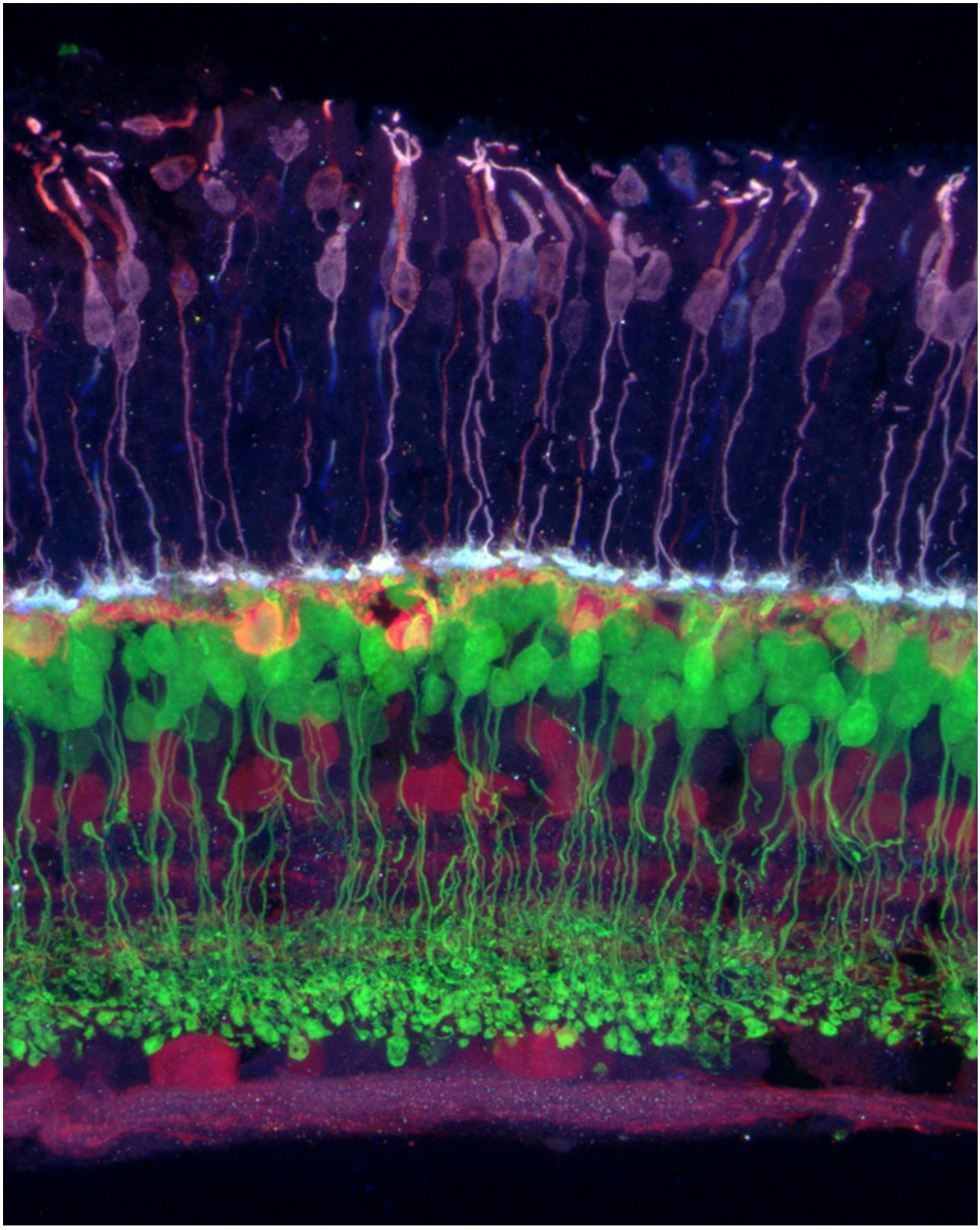
Sejnowski, who earlier this year became an advisor on the Obama administration's new BRAIN Initiative, a $100 million research effort to develop new techniques for studying the brain.

The easiest way for the brain to run the Boltzmann algorithm, he said, is to switch from beefing synapses up during the day to whittling them down during the night. Giulio Tononi, head of the Center for Sleep and Consciousness at the University of Wisconsin-Madison, has found that gene expression inside synapses changes in a way that supports this hypothesis: Genes involved in synaptic growth are more active during the day, and those involved in synaptic pruning are more active during sleep.

Alternatively, "the baseline could be computed during sleep and changes made relative to it during the day," Sejnowski said. His lab is building detailed computer models of synapses and the networks they sustain in order to determine how they collect firing statistics during wakefulness and sleep and when they change synaptic strengths to reflect the difference.

Brain Complications

A Boltzmann-like algorithm may be only one of many that the brain employs to tweak its synapses. In the 1990s, several independent groups developed a theoretical model of how the visual system efficiently encodes the flood of information striking the retina. The theory held that a process similar to image compression called "sparse coding" took place in the lowest layers of the visual cortex, making later stages of the visual system more efficient.



This image of the retina, in which different cell types are stained with different colors, highlights its layered structure. Color-sensitive cone cells (purple) connect to horizontal cells (orange), which connect to bipolar cells (green), and those connect to amacrine and ganglion cells (magenta).

The model's predictions are gradually passing more and more stringent experimental tests. In a paper published in PLOS Computational Biology in May, computational neuroscientists in the United Kingdom and Australia found that when neural networks using an algorithm for sparse coding called Products of Experts, invented by Hinton in 2002, are exposed to the same abnormal visual data as live cats (for example, the cats and neural networks both see only striped images), their neurons develop almost exactly the same abnormalities.

"By the time the information gets to the visual cortex, we think the brain is representing it as a sparse code," said Bruno Olshausen, a computational neuroscientist and director of the Redwood Center for Theoretical Neuroscience at the University of California-Berkeley, who helped develop the theory of sparse coding. "So it's like you have a Boltzmann machine sitting there in the back of your head trying to learn the relationships between the elements of the sparse code."

Olshausen and his research team recently used neural network models of higher layers of the visual cortex to show how brains are able to create stable perceptions of visual inputs in spite of image motion. In another recent study, they found that neuron firing activity throughout the visual cortex of cats watching a black-and-white movie was well described by a Boltzmann machine.

One potential application of that work is in the building of neural prosthesis, such as an artificial retina. With an understanding of "the formatting of information in the brain, you would know how to stimulate the brain to make someone think they are seeing an image," Olshausen said.

Sejnowski says understanding the algorithms by which synapses grow and shrink will enable researchers to alter them to study how network functions break down. "Then you can compare it to known problems that humans have," he said. "Almost all mental disorders can be traced to problems at synapses. So if we can understand synapses a little bit better, we'll be able to understand the normal function of the brain, how it processes information, how it learns, and what goes wrong when you have, say, schizophrenia."

The neural network approach to understanding the brain contrasts sharply with that of the Human Brain Project, Swiss neuroscientist Henry Markram's much-hyped plan to create a precise simulation of a human brain using a supercomputer. Unlike Hinton's approach of starting with a highly simplified model and gradually making it more complex, Markram wants to include as much detail as possible from the start, down to individual molecules, in hopes that full functionality and consciousness will emerge.

The project received $1.3 billion in funding from the European Commission in January, but Hinton thinks the mega-simulation will fail, mired by too many moving parts that no one yet understands. (Markram did not respond to requests for comment.)

More generally, Hinton doesn't think the workings of the brain can be deduced solely from the details of brain imaging studies; instead, these data should be used to build and refine algorithms. "You have to be thinking theoretically and exploring the space of learning algorithms to come up with a theory like" the Boltzmann machine, he said. For Hinton, the next step is to develop algorithms for training even more brainlike neural networks, such as ones that have synapses connecting neurons within, not just between, layers. "A major goal is to understand what you gain computationally by having more complicated computation at each stage," he said.

The hypothesis is that more interconnectedness enables stronger feedback loops, which, according to Olshausen, are probably how the brain achieves "perceptual filling in," where higher layers make inferences about what lower layers are sensing based on partial information. "That's intimately connected to consciousness," he said.

The human brain, of course, remains much more complicated than any of the models; it is larger, denser, more efficient, more interconnected, has more complex neurons — and juggles several algorithms simultaneously. Olshausen has estimated that we understand only 15 percent of the activity in the visual cortex. Although the models are making progress, neuroscience is still "a bit like physics before Newton," he said. Still, he is confident that the process of building on these algorithms may one day explain the ultimate riddle of the brain — how sensory data gets

transformed into a subjective awareness of reality. Consciousness, Olshausen said, “is something that emerges out of a really, really complicated Boltzmann machine.”

This article was reprinted on ScientificAmerican.com.