What Intelligent Machines Need to Learn From the Neocortex

Machines won’t become intelligent unless they incorporate certain features of the human brain. Here are three of them

By **JEFF HAWKINS** Posted 2 Jun 2017 | 15:00 GMT

Computers have transformed work and play, transportation and medicine, entertainment and sports. Yet for all their power, these machines still cannot perform simple tasks that a child can do, such as navigating an unknown room or using a pencil.

The solution is finally coming within reach. It will emerge from the intersection of two major pursuits: the reverse engineering of the brain and the burgeoning field of artificial intelligence. Over the next 20 years, these two pursuits will combine to usher in a new epoch of intelligent machines.

Why do we need to know how the brain works to build intelligent machines? Although machine-learning techniques such as deep neural networks have recently made impressive gains, they are still a world away from being intelligent, from being able to understand and act in the world the way that we do. The only example of intelligence, of the ability to learn from the world, to plan and to execute, is the brain. Therefore, we must understand the principles underlying human intelligence and use them to guide us in the development of truly intelligent machines.

At my company, Numenta (http://numenta.com/), in Redwood City, Calif., we study the neocortex, the brain’s largest component and the one most responsible for intelligence. Our goal is to understand how it works and to identify the underlying principles of human cognition. In recent years, we have made significant strides in our work, and we have identified several features of biological intelligence that we believe will need to be incorporated into future thinking machines.

To understand these principles, we must start with some basic biology. The human brain is similar to a reptile's brain. Each has a spinal cord, which controls reflex behaviors; a brain stem, which controls autonomic behaviors such as breathing and heart rate; and a midbrain, which controls emotions and basic behaviors. But humans, indeed all mammals, have something reptiles don't: a neocortex. (http://dev.biologists.org/content/141/1/11)

CAN WE COPY THE BRAIN?

Section 2: The Mechanics of the Mind

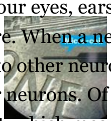
(/static/special-report-can-we-copy-the-brain)

The neocortex is a deeply folded sheet some 2 millimeters thick that, if laid out flat, would be about as big as a large dinner napkin. In humans, it takes up about 75 percent of the brain's volume. This is the part that makes us smart.

At birth, the neocortex knows almost nothing; it learns through experience. Everything we learn about the world—driving a car, operating a coffee machine, and the thousands of other things we interact with every day—is stored in the neocortex. It learns what these objects are, where they are in the world, and how they behave. The neocortex also generates motor commands, so when you make a meal or write software it is the neocortex controlling these behaviors. Language, too, is created and understood by the neocortex.

The neocortex, like all of the brain and nervous system, is made up of cells called neurons. (http://www.brainfacts.org/brain-basics/neuroanatomy/articles/2012/the-neuron/) Thus, to understand how the brain works, you need to start with the neuron. Your neocortex has about 30 billion of them. A typical neuron has a single tail-like axon (https://www.sciencedaily.com/terms/axon.htm) and several treelike extensions called dendrites. (http://www.cerebromente.org.br/n07/fundamentos/neuron/parts_i.htm) If you think of the neuron as a kind of signaling system, the axon is the transmitter and the dendrites are the receivers. Along the branches of the dendrites lie some 5,000 to 10,000 synapses, each of which connects to counterparts on thousands of other neurons. There are thus more than 100 trillion (https://decodethemind.wordpress.com/2010/09/10/more-brain-connections-than-stars-in-the-universe-no-not-even-close/) synaptic connections.

Your experience of the world around you—recognizing a friend's face, enjoying a piece of music, holding a bar of soap in your hand—is the result of input from your eyes, ears, and other sensory organs traveling to your neocortex and causing groups of neurons to fire. When a neuron fires, an electrochemical spike travels down the neuron's axon and crosses synapses to other neurons. If a receiving neuron gets enough input, it might then fire in response and activate other neurons. Of the 30 billion neurons in the neocortex, 1 or 2 percent are firing at any given instant, which means that many millions of neurons will be active at any point in time. The set of active neurons changes as you move and interact with the world. Your perception of the world, what you might consider your conscious experience, is determined by the constantly changing pattern of active neurons.



The neocortex stores these patterns primarily by forming new synapses. This storage enables you to recognize faces and places when you see them again, and also recall them from your memory. For example, when you think of your friend's face, a pattern of neural firing occurs in the neocortex that is similar to the one that occurs when you are actually seeing your friend's face.

Remarkably, the neocortex is both complex and simple at the same time. It is complex because it is divided into dozens of regions, each responsible for different cognitive functions. Within each region

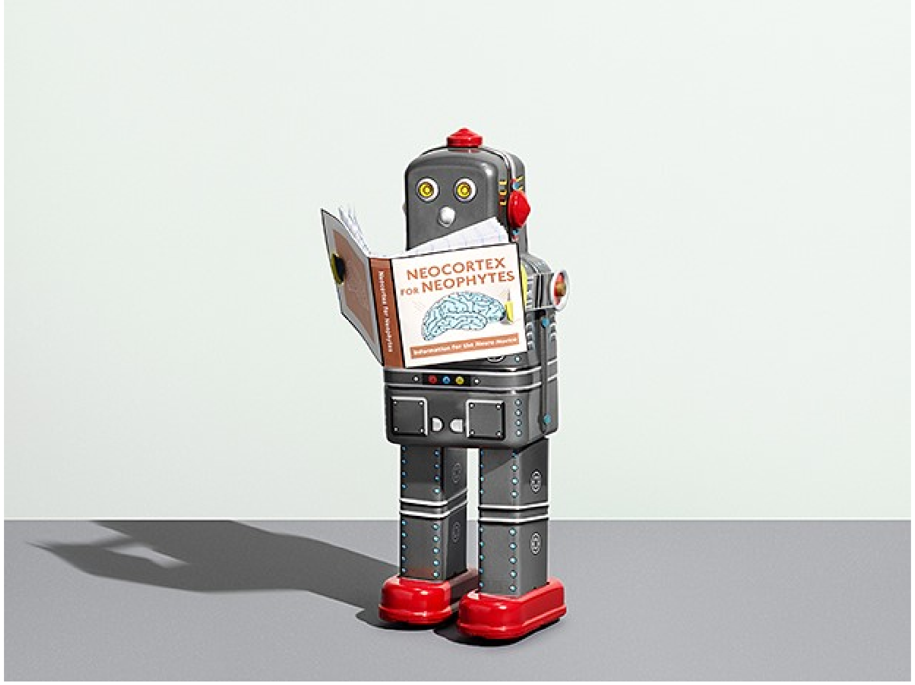


Photo: Dan Saelinger

there are multiple layers of neurons, as well as dozens of neuron types, and the neurons are connected in intricate patterns.

The neocortex is also simple because the details in every region are nearly identical. Through evolution, a single algorithm developed that can be applied to all the things a neocortex does. The existence of such a universal algorithm is exciting because if we can figure out what that algorithm is, we can get at the heart of what it means to be intelligent, and incorporate that knowledge into future machines.

But isn't that what AI is already doing? Isn't most of AI built on "neural networks (http://neuralnetworksanddeeplearning.com/)" similar to those in the brain? Not really. While it is true that today's AI techniques reference neuroscience, they use an overly simplified neuron model, one that omits essential features of real neurons, and they are connected in ways that do not reflect the reality of our brain's complex architecture. These differences are many, and they matter. They are why AI today may be good at labeling images or recognizing spoken words but is not able to reason, plan, and act in creative ways.

Our recent advances in understanding how the neocortex works give us insights into how future thinking machines will work. I am going to describe three aspects of biological intelligence that are essential, but largely missing from today's AI. They are learning by rewiring, sparse representations, and embodiment, which refers to the use of movement to learn about the world.

Learning by rewiring: Brains exhibit some remarkable learning properties. First, we learn quickly. A few glances or a few touches with the fingers are often sufficient to learn something new. Second, learning is incremental. We can learn something new without retraining the entire brain or forgetting what we learned before. Third, brains learn continuously. As we move around the world, planning and acting, we never stop learning. Fast, incremental, and continuous learning are essential ingredients that enable intelligent systems to adapt to a changing world. The neuron is responsible for learning, and the complexities of real neurons are what make it a powerful learning machine.

In recent years, neuroscientists have learned some remarkable things about the dendrite. One is that each of its branches acts as a set of pattern detectors. It turns out that just 15 to 20 active synapses on a branch are sufficient to recognize a pattern of activity in a large population of neurons. Therefore, a single neuron can recognize hundreds of distinct patterns. Some of these recognized patterns cause the neuron to become active, but others change the internal state of the cell and act as a prediction of future activity.

Neuroscientists used to believe that learning occurred solely by modifying the effectiveness of existing synapses so that when an input arrived at a synapse it would either be more likely or less likely to make the cell fire. However, we now know that most learning results from growing new synapses (https://www.sciencedaily.com/releases/2013/10/131010205325.htm) between cells—by "rewiring" the brain. Up to 40 percent of the synapses on a neuron are replaced with new ones every day. New synapses result in new patterns of connections among neurons, and therefore new memories. Because the branches of a dendrite are mostly independent, when a neuron learns to recognize a new pattern on one of its dendrites, it doesn't interfere with what the neuron has already learned on other dendrites.

This is why we can learn new things without interfering with old memories and why we don't have to retrain the brain every time we learn something new. Today's neural networks don't have these properties.

Neuron Mimicry Explained

▶

Intelligent machines don’t have to model all the complexity of biological neurons, but the capabilities enabled by dendrites and learning by rewiring are essential. These capabilities will need to be in future AI systems.

Sparse representations: Brains and computers represent information quite differently. In a computer’s memory, all combinations of 1s and 0s are potentially valid, so if you change one bit it will typically result in an entirely different meaning, in much the same way that changing the letter *i* to *a* in the word *fire* results in an unrelated word, *fare*. Such a representation is therefore brittle.

Brains, on the other hand, use what’s called sparse distributed representations (http://www.cortical.io/technology_representations.html), or SDRs. They’re called sparse because relatively few neurons are fully active at any given time. Which neurons are active changes moment to moment as you move and think, but the percentage is always small. If we think of each neuron as a bit, then to represent a piece of information the brain uses thousands of bits (many more than the 8 to 64 used in computers), but only a small percentage of the bits are 1 at any time; the rest are 0.

Let's say you want to represent the concept of "cat" using an SDR. You might use 10,000 neurons of which 100 are active. Each of the active neurons represents some aspect of a cat, such as "pet," or "furry," or "clawed." If a few neurons die, or a few extra neurons become active, the new SDR will still be a good representation of "cat" because most of the active neurons are still the same. SDRs are thus not brittle but inherently robust to errors and noise. When we build silicon versions of the brain, they will be intrinsically fault tolerant.

There are two properties of SDRs I want to mention. One, the overlap property, makes it easy to see how two things are similar or different in meaning. Imagine you have one SDR representing "cat" and another representing "bird." Both the "cat" and "bird" SDR would have the same active neurons representing "pet" and "clawed," but they wouldn't share the neuron for "furry." This example is simplified, but the overlap property is important because it makes it immediately clear to the brain how the two objects are similar or different. This property confers the power to generalize, a capability lacking in computers.

The second, the union property, allows the brain to represent multiple ideas simultaneously. Imagine I see an animal moving in the bushes, but I got only a glimpse, so I can't be sure of what I saw. It might be a cat, a dog, or a monkey. Because SDRs are sparse, a population of neurons can activate all three SDRs at the same time and not get confused, because the SDRs will not interfere with one another. The ability of neurons to constantly form unions of SDRs makes them very good at handling uncertainty.

Such properties of SDRs are fundamental to understanding, thinking, and planning in the brain. We can't build intelligent machines without embracing SDRs.

Embodiment: The neocortex receives input from the sensory organs. Every time we move our eyes, limbs, or body, the sensory inputs change. This constantly changing input is the primary mechanism the brain uses to learn about the world. Imagine I present you with an object you have never seen before. For the sake of discussion, let's say it's a stapler. How would you learn about the new object? You might walk around the stapler, looking at it from different angles. You might pick it up, run your fingers over it, and rotate it in your hands. You then might push and pull on it to see how it behaves. Through this interactive process, you learn the shape of the stapler, what it feels like, what it looks like, and how it behaves. You make a movement, see how the inputs change, make another movement, see how the inputs change again, and so on. Learning through movement is the brain's primary means for learning. It will be a central component of all truly intelligent systems.

This is not to say that an intelligent machine needs a physical body, only that it can change what it senses by moving. For example, a virtual AI machine could "move" through the Web by following links and opening files. It could learn the structure of a virtual world through virtual movements, analogous to what we do when walking through a building.

This brings us to an important discovery we made at Numenta last year. In the neocortex, sensory input is processed in a hierarchy of regions. As sensory input passes from one level of the hierarchy to another, more complex features are extracted, until at some point an object can be recognized. Deep-learning networks also use hierarchies, but they often require 100 levels of processing to recognize an image, whereas the neocortex achieves the same result with just four levels. Deep-learning networks also require millions of training patterns, while the neocortex can learn new objects with just a few movements and sensations. The brain is doing something fundamentally different than a typical artificial neural network, but what?

Hermann von Helmholtz (https://plato.stanford.edu/entries/hermann-helmholtz/), the 19th-century German scientist, was one of the first people to suggest an answer. He observed that, although our eyes move three to four times a second, our visual perception is stable. He deduced that the brain must take account of how the eyes are moving; otherwise it would appear as if the world were wildly jumping about. Similarly, as you touch something, it would be confusing if the brain processed only the tactile input and didn't know how your fingers were moving at the same time. This principle of combining movement with changing sensations is called sensorimotor integration. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3154729/) How and where sensorimotor integration occurs in the brain is mostly a mystery.

Our discovery is that sensorimotor integration occurs in every region of the neocortex. It is not a separate step but an integral part of all sensory processing. Sensorimotor integration is a key part of the "intelligence algorithm" of the neocortex. We at Numenta have a theory and a model of exactly how neurons do this, one that maps well onto the complex anatomy seen in every neocortical region.

What are the implications of this discovery for machine intelligence? Consider two types of files you might find on a computer. One is an image file produced by a camera, and the other is a computer-aided design file produced by a program such as Autodesk. An image file represents a two-dimensional array of visual features. A CAD file also represents a set of features, but each feature is assigned a location in three-dimensional space. A CAD file models complete objects, not how the object appears from one perspective. With a CAD file, you can predict what an object will look like from any direction and determine how an object will interact with other 3D objects. You can't do these with an image file. Our discovery is that every region of the neocortex learns 3D models of objects much like a CAD program. Every time your body moves, the neocortex takes the current motor command, converts it into a location in the object's reference frame, and then combines the location with the sensory input to learn 3D models of the world.

In hindsight, this observation makes sense. Intelligent systems need to learn multidimensional models of the world. Sensorimotor integration doesn't occur in a few places in the brain; it is a core principle of brain function, part of the intelligence algorithm. Intelligent machines also must work this way.

These three fundamental attributes of the neocortex—learning by rewiring, sparse distributed representations, and sensorimotor integration—will be cornerstones of machine intelligence. Future thinking machines can ignore many aspects of biology, but not these three. Undoubtedly, there will be other discoveries about neurobiology that reveal other aspects of cognition that will need to be incorporated into such machines in the future, but we can get started with what we know today.

From the earliest days of AI, critics dismissed the idea of trying to emulate human brains, often with the refrain that "airplanes don't flap their wings." In reality, Wilbur and Orville Wright studied birds in detail. To create lift, they studied bird-wing shapes and tested them in a wind tunnel. For propulsion, they went with a nonavian solution: propeller and motor. To control flight, they observed that birds twist their wings to bank and use their tails to maintain altitude during the turn. So that's what they did, too. Airplanes still use this method today, although we twist only the tail edge of the wings. In short, the Wright brothers studied birds and then chose which elements of bird flight were essential for human flight and which could be ignored. That's what we'll do to build thinking machines.

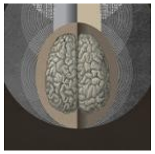
As I consider the future, I worry that we are not aiming high enough. While it is exciting for today's computers to classify images and recognize spoken queries, we are not close to building truly intelligent machines. I believe it is vitally important that we do so. The future success and even survival of humanity may depend on it. For example, if we are ever to inhabit other planets, we will need machines to act on our behalf, travel through space, build structures, mine resources, and independently solve complex problems in environments where humans cannot survive. Here on Earth, we face challenges related to disease, climate, and energy. Intelligent machines can help. For example, it should be possible to design intelligent machines that sense and act at the molecular scale. These machines would think about protein folding and gene expression in the same way you and I think about computers and staplers. They could think and act a million times as fast as a human. Such machines could cure diseases and keep our world habitable.

In the 1940s, the pioneers of the computing age sensed that computing was going to be big and beneficial, and that it would likely transform human society. But they could not predict exactly how computers would change our lives. Similarly, we can be confident that truly intelligent machines will transform our world for the better, even if today we can't predict exactly how. In 20 years, we will look back and see this as the time when advances in brain theory and machine learning started the era of true machine intelligence.

About the Author

Jeff Hawkins is the cofounder of Numenta (http://numenta.com/), a Redwood City, Calif., company that aims to reverse engineer the neocortex.

SPECIAL REPORT: CAN WE COPY THE BRAIN?
(/static/special-report-can-we-copy-the-brain)



PREVIOUS
The Brain as Computer: Bad at Math, Good at Everything Else

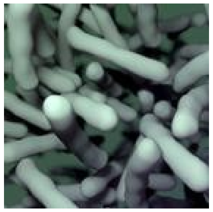
NEXT
How to Digitize a Rat Brain
(/video/biomedical/imaging/how-to-digitize-a-rat-brain)

(/computing/hardware/the-brain-as-computer-bad-at-math-good-at-everything-else)

Recommended For You

Startup Neurable Unveils the World's First Brain-Controlled VR Game

(/the-human-os/biomedical/bionics/brainy-startup-neurable-unveils-the-worlds-first-braincontrolled-vr-game)

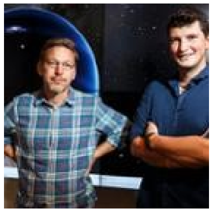


AI Makes Anthrax Bioterror Detection Easier

(/the-human-os/biomedical/imaging/ai-makes-anthrax-bioterror-detection-easier)

Teenage Whiz Kid Invents an AI System to Diagnose Her Grandfather's Eye Disease

(/the-human-os/biomedical/diagnostics/teenage-whiz-kid-invents-an-ai-system-to-diagnose-her-grandfathers-eye-disease)

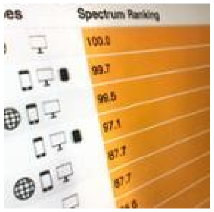


Is There a Giant Planet Lurking Beyond Pluto?

(/aerospace/satellites/is-there-a-giant-planet-lurking-beyond-pluto)

Smart Contact Lenses and Eye Implants Will Give Doctors Medical Insights

(/biomedical/devices/smart-contact-lenses-and-eye-implants-will-give-doctors-medical-insights)



The 2017 Top Programming Languages

(/computing/software/the-2017-top-programming-languages)