Biologically Driven Artificial Intelligence

Kjell J. Hole, Simula UiB

Subutai Ahmad, Numenta

Artificial Intelligence (AI) based on the computational principles of the brain can overcome current shortcomings and lead to human-level AI.

Artificial intelligence (AI) is the study of techniques that allow computers to learn, reason, and act to achieve goals. Machine learning is an essential type of AI that permits machines to learn from big data sets without being explicitly programmed. At the time of this writing, the international AI community is focusing on a type of machine learning called *deep learning*, a family of statistical techniques for classifying patterns using artificial neural networks (ANNs) with many layers.[1] Although deep learning has led to substantial advances in image and speech recognition, language translation, and game playing, we argue that the AI community should focus less on methods using nonbiological ANNs and more on the computational principles of the brain to create human-level or general AI with a performance close to humans at most cognitive tasks of interest.[2,3]

Digital Object Identifier 10.1109/MC.2019.2917455
Date of publication: 30 July 2019

HUMAN VERSUS ARTIFICIAL INTELLIGENCE

Human intelligence is the brain's ability to predict sensations and events, learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate the environment. The outermost layer of the brain, the neocortex, controls cognitive functions (see "The Neocortex"). Human cognition depends on the tight connections between the brain and the body. The neocortex not only integrates sensory processing and generates motor commands, but it also uses the same sensorimotor mechanisms for high-level cognition. In other words, the structures and functions of the brain and the body form the human thought processes.

Current AI consists of computer programs that process input data from the environment and generate output data. AI researchers have mostly ignored studies of the neocortex and instead focused on mathematical and logical methods to develop AI. Because the brain is a product of evolution where newer components, including the neocortex, must rely on the functionality of the older parts, there is a widespread belief that it is possible to develop general AI by starting from scratch. AI researchers have also emphasized the limitations of the brain: the skull restricts the brain's physical size, access to energy from the body is limited, and the speed of the brain's neural circuits is slow compared to modern computers.

 0018-9162/19©2019IEEE

EDITOR ERIK P. DEBENEDICTIS
Sandia National Laboratories;
epdeben@sandia.gov

CURRENT LIMITATIONS OF AI

Even though it may be theoretically possible to create superior AI systems that are not biologically inspired, we are far from the goal of creating AI at the level of human intelligence. Most AI learning algorithms, particularly deep learning algorithms, are greedy, brittle, rigid, and opaque.[2] The algorithms are

- greedy because they demand big data sets to learn
- brittle because they frequently fail when confronted with a mildly different scenario than that in the training set
- rigid because they cannot keep adapting after initial training
- opaque because the internal representations make it challenging to interpret their decisions.

Although these shortcomings are all serious, the core problem is that all AI systems are shallow because they lack abstract reasoning abilities and possess no common sense about the world.

The most popular AI methods today use ANNs. The brain was the original inspiration for ANNs, but this is no longer the case; the knowledge behind the ANNs' brain cells is outdated. We now know that biological neurons have multiple physical states and biological networks have far more sophisticated functionality than those in ANNs. Furthermore, most AI systems require separate, offline training periods, whereas learning in the neocortex occurs continuously and in real time using data streaming in from all our senses. Finally, sensorimotor processing is not deeply integrated into most AI methods. It is therefore unclear whether current AI approaches can lead to general human-level intelligence.

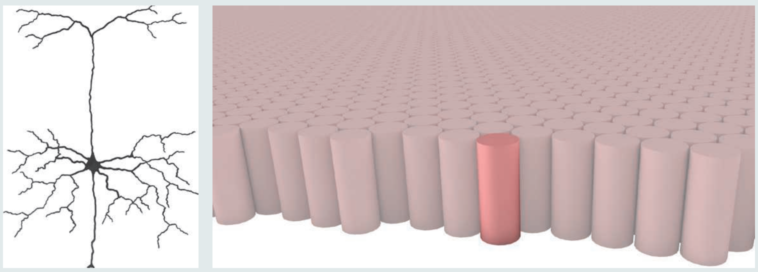
BIOLOGICALLY BASED AI

The limited progress toward general AI after decades of research strongly

THE NEOCORTEX

The neocortex is an intensely folded sheet with a thickness of approximately 2.5 mm. The folds substantially increase its surface area. When laid out flat, the neocortex is the size of a formal dinner napkin. It constitutes roughly 75% of the brain's volume and contains 30 billion cells called *neurons* [see Figure S1(a)]. A typical neuron has one tail-like axon and several tree-like extensions called *dendrites*. When a cell fires, an electrochemical pulse or spike travels down the axon to its terminals. A signal jumps from an axon terminal to the receptors on a dendrite of another neuron. The axon terminal, the receptors, and the cleft between them constitute a synapse (not shown). The axon terminal releases neurotransmitters into the synaptic cleft to signal the dendrite. The neuron is thus a signaling system in which the axon is the transmitter, the dendrites are the receivers, and the synapses are the connectors between them. A neuron has between 5,000 and 20,000 synapses.

The neocortex consists of cortical columns as depicted in Figure S1(b). The columns have six layers, five of which contain neurons. The thickness of a layer varies throughout the neocortex, but the basic structure is similar. As a result, the neocortex has both a vertical columnar and horizontal laminar organization. Cortical columns are grouped into regions connected through bundles of nerve fibers and organized in hierarchies. The regions realizing vision, language, and touch all have highly similar structures and thus operate according to the same principles. The sensory input determines a region's purpose. Some regions receive inputs directly from the sensory organs, while others receive them from other regions. The neocortex generates body movements to change the sensory inputs and learn quickly about the world. Learning occurs by growing and removing synapses, that is, by changing the network itself.



(a) (b)

FIGURE S1. (a) A biological neuron in the neocortex. The axon is the lowermost vertical connection. (b) A flattened version of the neocortex consisting of cortical columns, all with a similar structure.

indicates that few paths lead to human-level intelligence and many paths lead nowhere. Not surprisingly, without guidance, it is challenging to discover the right approach in a vast space of algorithms containing very few solutions. Instead, the community should focus on the only example we have of intelligence, namely, the brain and, especially, the neocortex.

There has been much effort in neuroscience to reverse engineer mammalian brains, especially to understand what the neocortex does and how it works. Substantial reverse engineering efforts include the U.S. government program Machine Intelligence from Cortical Networks, the Swiss Blue Brain Project, and the Human Brain Project funded mainly by the European Union. The research lab Numenta, led by Jeff Hawkins, has shared a theoretical framework, called *Hierarchical Temporal Memory* (*HTM*), that describes the computational principles of the neocortex.[4–6] Numenta has made a sustained effort to make HTM understandable to people without a neuroscience background (see research papers and tutorial videos at numenta.com).

One of the findings of neuroscience is that the neocortex consists of a repeating circuit, known as a *cortical column*, that creates our perceptions, language, and high-level thoughts. HTM theory models several components of this circuit that, together, have led to a new functional interpretation of the neocortex. We summarize these findings and the differences between biologically based frameworks and ANN-based systems in the following sections.[4–6]

Realistic neuron model

AI models neurons as a simple function operating on a single set of connections. Biological neurons, modeled by the HTM neuron, have segregated feedforward, lateral, and feedback connections that detect multiple independent patterns of neuronal activity. Whereas the ANN neuron is either active or inactive, the HTM neuron has more states. The HTM neuron goes into a predictive state when it is soon likely to become active based on lateral and feedback context. The ability to recognize contextual patterns and predict future states form the basis for a predictive sequence memory that allows HTM to process multiple hypotheses about the world simultaneously.

Sparse data representations

An ANN uses dense scalar vectors to represent data. Such representations are vulnerable to errors and noise, and dense vector-matrix multiplications consume significant power. The brain uses sparse data representations where only a small percentage of the neurons is active, and each neuron is only connected to a fraction of other neurons. Sparse representations allow HTM to represent and process patterns of neuronal activity in a brain-like manner that is robust to changes and faults and consume significantly less power.

Sensorimotor integration

The brain models objects and environments by integrating sensations and movement-derived location signals. These object models are continuously updated as sensors move to create movement-invariant models. HTM uses such models to make predictions, detect anomalies, and thus learn about the world. It is unclear how to achieve the same results with ANNs.

Continuous online learning

To train most ANN systems, we must first assemble big training sets and then run learning algorithms offline. HTM learning occurs in real time using streaming data from sensors. Learning in the neocortex, and in HTM, dynamically rewires the connections between neurons to record information. Since the Hebbian-like learning rules are local to each HTM neuron, learning in the neocortex can scale to train massive systems.

The Thousand Brains Theory of Intelligence

Connectivity between cortical columns is strongly nonhierarchical, with numerous lateral and feedback pathways. HTM theory states that instead of learning one big model of the world, each cortical column learns a separate movement-invariant model. Communication between cortical columns cuts across hierarchy and sensory modalities and rapidly resolves uncertainty. The neocortex thus learns thousands of models that operate in a massively parallel fashion. This thousand brains theory of intelligence is built on a single universal learning algorithm embodied by the cortical column. The algorithm learns objects, behaviors, and concepts by representing compositional structures, learning through movement, and integrating knowledge across different senses. There is no comparable theory based on ANNs.

FURTHER RESEARCH

Although deep learning can outperform humans in isolated domains characterized by fixed rules and little need for contextual knowledge,[1] we require novel ideas to create general AI. The biological path toward human-level AI, particularly the principles discussed previously, deserves more attention from the AI community because biologically constrained AI avoids the limitations of the approaches based on ANNs. More computer scientists should study neuroscience and use the computational principles of the brain to develop AI. Students and budding researchers with a desire to contribute to the development of AI should also study neuroscience. Finally, we need increased cooperation with neuroscientists, given how hard it is to select relevant results from the vast neuroscience literature.

There is a rich road map for future research into biologically constrained AI. Additional reverse engineering is required to understand the organization of the neocortical regions and how they collaborate to create intelligent behavior. It is also necessary to reverse engineer other parts of the brain, such as the thalamus, which is intimately involved in the communication between the neocortical regions.

It will take effort to follow the biological path to general AI, but repeated improvements along the way could gradually change society for the better. Initially, we could build specialized brains to improve cybersecurity or solve difficult problems in mathematics, physics, or medicine. Eventually, we could create autonomous intelligent systems to carry out tedious or dangerous work. ▣

REFERENCES

1. T. J. Sejnowski, *The Deep Learning Revolution*. Cambridge, MA: MIT Press, 2018.
2. G. Marcus, Deep learning: A critical appraisal, 2018. [Online]. Available: https://arxiv.org/abs/1801.00631.
3. D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.
4. J. Hawkins and S. Ahmad, "Why neurons have thousands of synapses, a theory of sequence memory in neocortex," *Front. Neural Circuits*, vol. 10, March 2016. [Online]. Available: https://doi.org/10.3389/fncir.2016.00023
5. J. Hawkins, S. Ahmad, and Y. Cui, "A theory of how columns in the neocortex enable learning the structure of the world," *Front. Neural Circuits*, vol. 11, Oct. 2017. [Online]. Available: https://doi.org/10.3389/fncir.2017.00081
6. J. Hawkins, M. Lewis, M. Klukas, S. Purdy, and S. Ahmad, "A framework for intelligence and cortical function based on grid cells in the neocortex," *Front. Neural Circuits*, vol. 12, Jan. 2019. [Online]. Available: https://doi.org/10.3389/fncir.2018.00121

KJELL J. HOLE is the director of the research center Simula UiB, Bergen, Norway. Contact him at hole@simula.no.

SUBUTAI AHMAD is the vice president of research at Numenta, Redwood City, California. Contact him at sahmad@numenta.com.

Digital Object Identifier 10.1109/MC.2019.2926582