Brains Are Automata: This is Not Just AI and Robotics

Juyang Weng

Abstract—**Experts have said many times that we humans still do not know how the brain works. However, this is no longer true. A brain automata theory appeared in 2011 with the support of a series of ground-breaking computer simulation results [18]. The full and formal mathematical proof of the theory is now available [20]. It has been formally proved that all grounded brains are also Turing Machines — a theoretical model for modern computers (Von Neumann computers). This new theory has unified natural intelligence and artificial intelligence. More specifically, it unified the two schools in Artificial Intelligence (AI) — the symbolic school and the connectionist school. Knowledge of both schools contributed to the brain automata theory but the brain automata are like none of them.**

Index Terms—**Brains, Turing Machines, neural networks, deep learning**

Although there has been much work on studying brains, there has never been a brain-scale computational model like the brain automata model. Unlike traditional Turing Machines well studied in computer science, brain Turing Machines belong to a new kind — they are grounded, emergent, natural, incremental, attentive, and motivated. Unlike traditional neural networks, brain networks can not only do pattern recognition, but also abstract new concepts from examples and automatically program themselves for general purposes in the physical world. Traditional Turing Machines cannot do the same.

I. PROTOTYPES

Therefore, the time to make brain-like intelligent devices seems to have arrived, as the Where-What Networks (WWN) have demonstrated progressively as prototypes, from WWN-1 through WWN-9. Enjoying the popularity of the current inexpensive sensors, user-interfaces, and mobile computing devices, Dr. John Weng would like to find a mobile computing partner who is interested in producing a line of brain-like personal devices. He predicts that this consumer success will enable the partner to reshape the current trend of wearable computing and mobile computing. For example, wearable computers have a little personal "brain" that assists how the user moves around (walker safety) and drives around (driver safety). This user can be a child, a drinker, a blind, or an elderly. For example, the little silicon "brain" (1) warns when the user crosses a street, (2) serves as a "vision" pen for a blind, (3) prompts to slow down the car when a child runs towards the street.

J. Weng is with the Department of Computer Science and Engineering, the Cognitive Science Program, and the Neuroscience Program, Michigan State University, East Lansing, MI, 48824 USA e-mail: (see http://www.cse.msu.edu/~weng/). During mid-April 2015, the material in this article was made available as an executive summary to MSU administrators and some industrial executives.

From WWN-1 to WWN-9, the added advances are, respectively, WWN-1: from location to type (i.e., recognition) and from type to location (i.e., detection) by the same network; WWN-2: free-viewing: location and type of a learned object from natural cluttered scenes (i.e., detection and recognition simultaneously); WWN-3: dealing with multiple learned objects in natural cluttered sciences (i.e., detection and recognition are not unique); WWN-4: showing a static cascade of processing modules (deep learning) is not as good as dynamically emergent network of processing modules (not a cascade); WWN-5: dealing with different scales of objects; WWN-6: added synapse maintenance for neurons to automatically segment objects from clustered scenes; WWN-7: learn different scales of the same object (e.g., nose, eyes-and-nose, and face) while the skull is fully closed during development; WWN-8: add multi-sensory (i.e., left and right cameras) integration so as to learn to predict 3-D shape and 3-D object type but such 3D information is from not only stereo parallax but also intensity distribution such as texture; WWN-9: learn abstract relation of abstract objects.

The WWNs have been tested on brain functions related to vision (e.g., stereo vision, object recognition, object detection, all in natural cluttered scenes), audition (sound and speech), autonomous navigation, language understanding and language production. The emphasis has been on demonstrating how a machine "baby" can become "alive" to fully automatically learn while its body interacting with the real physical world and generalize in new settings, instead of human-handcrafted adult static capabilities. For example, no hand-crafted computer program can enable a machine to navigate in a new city or learn a new language.

II. HARD TO BELIEVE

Historically, public acceptance of science was slow. For example, Charles Darwin waited about 20 years (from the 1830s to 1858) to publish his theory of evolution for fear of public reaction. About 20 years later (by the 1870s) the scientific community and much of the general public had accepted evolution as a fact. Of course, the debate on evolution still goes on today.

Is the public acceptance of science faster in modern days? Not necessarily so, even though we have now better and faster means to communicate. The primary reason is still the same but much more severe—the remaining open scientific problems are more complex and the required knowledge goes beyond a typical single person.

For instance, network-like brain computation — connectionist computation (e.g., J. McClelland & D. Rumelhart 1986

[13], [10]) — has been long doubted and ignored by industry. Deep convolutional networks appeared by at least 1980 (K. Fukushima [5]). Max-pooling technique for deep convolutional networks was published by 1992 (J. Weng et al. [21], [22]). However, Apple, Baidu, Google, Microsoft, Samsung, and other major related companies did not show considerable interest till after 2012. That is a delay of about 20 years.

III. Misled intuition

Many machine learning experts believed that deep learning was the reason behind the recent rise of industrial interest, as Google acquired DNNresearch March 13, 2013. I knew some of the other reasons. The two techniques above, deep convolutional network and max-pooling used by G. Hinton et al. 2012 [8] in their winning ImageNet contest, are not very difficult to understand. They are more than 20 years old. These techniques have already been proved obsolete by the discoveries of more fundamental and effective principles of the brain, seven of which are intuitively explained in the following section. January 26, 2014, Google announced that it had agreed to acquire artificial intelligence company DeepMind Technologies. It seems that Google is interested in brain related technology.

Industrial and academic interests have been keen on a combination of two things — easily understandable tests (e.g., G. Hinton et al. 2012 [8], congratulations!) and major companies being involved (e.g., Google, thanks!). We have read statements like "our results can be improved simply by waiting for faster GPUs and bigger datasets to become available" (G. Hinton et al. 2012 [8]). However, the newly known brain principles have told us that the ways to conduct such tests (e.g., ImageNet) will give only vanishing gains that do not lead to a human-like zero error rate, regardless how long the Moore's Law can continue and how many more static images are added to the training set. Why? All such tests used static images in which objects mix with the background. Such tests therefore prevent participating groups from seriously considering autonomous object segmentation (free of handcrafted object model). Through synapse maintenance (Y. Wang et al. 2011 [17]), neurons in a human brain automatically cut off inputs from background pixels if background pixels matched badly compared with object pixels. Our babies spend much more time in dynamic physical world than seeing static photos.

Our industry needs to learn more powerful brain mechanisms that go beyond conventional well-known, well-tested techniques. The following gives some examples:

(1) **Deep Learning Networks** (e.g., J. Weng et al. 1992 [21], Y. LeCun et al. 1998 [9], G. Hinton et al. 2012 [8]) are not only biologically implausible but also functionally weak. The brain uses a rich network of processing areas (e.g., Felleman & Van Essen 1991 [4]) where connections are almost always two-way, not a cascade of modules like the Deep Learning Networks. Such a Deep Learning Network is not able to conduct top-down attention in a cluttered scene (e.g., attention to location or type in J. Weng 2012 [19] or more complex object shape as reported in L. B. Smith et al. 2005 [14]).

(2) **Convolution** (e.g., J. Weng et al. 1992 [21], Y. LeCun et al. 1998 [9], G. Hinton et al. 2012 [8]) is not only biologically implausible, but also computationally weak. Why? All feature neurons in the brain carry not only sensory information but also motor information (e.g., Felleman & Van Essen 1991 [4]) so that later-processing neurons become less concrete and more abstract — which is impossible to accomplish using the shift-invariant convolution. Namely, convolution is always location-concrete (even using max-pulling) and never location-abstract.

(3) **Error back-propagation** in neural networks (e.g., Y. LeCun et al. 1998 [9], G. Hinton et al. 2012 [8]) is not only biologically implausible (e.g., a baby does not have error in his motors) but also damaging to long-term memory because of its lack of match-based competition for error-causality (such as those in the global-input SOM [7], the local-inputs LISSOM [11], and the spatiotemporally optimal LCA [23]). Even though the gradient vector identifies the neuron that can most reduce the current error, the current error is not the "business" of that neuron at all. Instead it should keep its own long-term memory unchanged. That is why error back-propagation is well known to be bad for incremental learning. That is why it requires research assistants to try many guesses of initial weights (i.e., using the test set as the training set!). Let us not be blinded by artificially low error rates.

The brain principles below state that every network must be successful in its "life".

IV. Brain principles

To understand the potential of the brain technology, the following brain principles are necessary.

In terms of brain computation, the circuits in your brain self-wire beautifully and precisely according to your real-time experience (the genome only regulates).

Traditionally, many domain experts think that computers and brain appear to use very different principles. Naturally emerging Turing Machine in Developmental Networks that has been mathematically proved (see J. Weng, Brain as an Emergent Finite Automaton: A Theory and Three Theorems, 2015 [20]) should change that intuition. The brain automata theory proposed the following seven brain principles:

1) **Autonomous development.** A developmental program (genome-like, task-nonspecific) regulates the autonomous development (i.e., Elman et al. [3], Weng et al. [24]) of a task-nonspecific brain-like network — Developmental Network [20]. The Developmental Network is of general-purpose—can learn any body-capable tasks, in principle. Not only pattern recognition.
2) **Automatic segmentation.** The brain's images are naturally sensed images of cluttered scenes where many objects mix. In traditional machine training (e.g., G. Hinton et al. 2012 [8]), each training image has a bounding box drawn around each object to learn, which is not the "luxury" that a human baby can afford. Neurons in the Developmental Network automatically determine which object to attend and learn.
3) **Sensory abstraction.** The brain attends to an object among many objects in a cluttered scene. A mature

brain's muscles are sensorily abstract. The mouth says "apple" correctly regardless where the apple is; the arm reaches correctly for an apple or toy regardless what object type it is. The brain can figure out declarative knowledge (e.g., abstract concepts such as location, type, scale, etc. of an apple) and non-declarative knowledge (e.g., driving a car or riding a bicycle). The brain does not just do global pattern classification.

4) **Incremental.** The brain's network learns incrementally — taking one-pair of sensory pattern (e.g., two retinal images) and motor pattern (e.g., muscle firing pattern) at a time to update the brain and discarding the sensory-motor pair immediately after. Namely, a real brain has only one pair of stereoscopic retinas which cannot store more than one pair of images. Batch learning (i.e., learn fully before test) is not scalable. Early learned skills help later learning.
5) **Motivated.** Avoid pain and seek pleasure to speed up learning — spend more memory resources on important events. Neurons estimate uncertainty and novelty to automatically and dynamically determine where to connect (e.g., cut off background pixels for object segmentation).
6) **Brain automata.** Each brain is an Emergent Turing Machine (Weng 2015 [20]). Every area in the brain network emerges (does not statically exist, see M. Sur et al. 2000 [15] and P. Voss 2013 [16]) using a unified area function whose feature development is nonlinear but free of local minima, contrary to engineering intuition — not convolution; not error back-propagation. The logic completeness of a brain (e.g., general purpose) is explained by a Universal Turing Machine in a Developmental Network.
7) **Optimal.** The brain's network is always optimal. It updates in real time to compute the maximum likelihood estimate of the sensory-motor space, conditioned on the limited computational resources (i.e., the size of the brain) and the limited learning experience in its life so far. Some neural transmitters differentially bias the weights for some events in lifetime — important events (e.g., pain) have more weights in the optimal representations. Every child must "group up" successfully. One should not use the test set as a training set as genetic algorithms do to "kill" all children except the single best. It is also too costly to try many networks and hand-tune network parameters.

V. Crossdisciplinary supports

Neuroscience and psychology have made many advances by providing experimental data (e.g., Felleman & Van Essen 1991 [4]). However, it has been well recognized that these disciplines are data-rich and theory-poor. The phenomena of brain circuits and brain behavior are extremely rich. Many researchers in these areas use only local tools (e.g., attracters that can only be attracted into local extrema) and consequently have been overwhelmed by the richness of brain phenomena. A fundamental reason is that they miss the guidance of the global automata theory of computer science. For example, X. Wang et al. 2013 [12] stated correctly that neurons of mixed selectivity were rarely analyzed but have widely observed. However, the mixed selectivity has already been well explained, as a special case, by the new Emergent Turing Machine in Developmental Networks in a theoretically complete way. The mixed selectivity of neurons are caused by emergent and beautiful brain circuits as a Turing Machine.

The brain automata model has predicted that brain circuits dynamically and precisely record the statistics of experience, roughly consistent with neural anatomy (e.g., Felleman & Van Essen 1991 [4]). In particular, the model predicted that shifting attention between 'humans' and 'vehicles' dramatically changes brain representation of all categories as reported by J. Gallant et al. 2013 [2]. It also predicts that human attention can regulate the activity of their neurons in the medial temporal lobe as reported by C. Koch et al. 2011 [1]. The place cells work of the 2014 Nobel Prize in Physiology or Medicine implies that neurons encode exclusively bottom-up information (place). The brain automata model challenges such a view: Neurons represent a combination of both bottom-up (e.g., place) and top-down context (e.g., goal) as reported by Koch et al. [1] and Gallant et al. [2].

VI. Require more than domain expertise

October 2011, a highly respected multi-disciplinary professor kindly wrote: I tell these students that they can work on brains and do good science, or work on robots and do good engineering. But if they try to do both at once, the result will be neither good science nor good engineering. However, the pessimistic view of the brain was no longer true even then.

The fundamental barrier is not whether a person works in the symbolic school or the connectionist school, but rather, the lack of knowledge in at least six disciplines — biology, neuroscience, psychology, computer science, electrical engineering, and mathematics. Therefore, the brain automata theory is not something that an expert can understand through a seminar, or even a single course. He must make up the knowledge that he lacks in all the six disciplines.

Unfortunately, the brain automata model implies that all neuroscientists are unable to understand the brain of their studies without a rigorous training in automata theory, neural networks, and mathematics. Likewise, all neural network modelers will not be able to understand how the brain works without a systematic training in the automata theory, biology, neuroscience, and psychology. For example, traditional models for nervous systems and neural networks focus on pattern recognition and do not have the capabilities of "rules" (e.g., "rulefully combining and recombining" stated by Stevan Harnad 1990 [6]). A human student can first learn rules in a class according to his own observations of what demonstrated in the class. Next, he applies the rules to his homework and real life. The brain automata theory deals with such "rule-like" capabilities.

VII. Why now?

No species on this planet has reached this stage other than human — human as a species largely understands how its own brain works. The brain automata theory, algorithm, and success of early prototypes have opened the door toward fundamental changes in the way humans live. The current mobile computing technology seems to be ready for computations in a small silicon brain.

Does this new knowledge stun our students, researchers and entrepreneurs or instead guide them to better spend their time and other resources?

REFERENCES

[1] M. Cerf, N. Thiruvengadam, F. Mormann, A. Kraskov, R. Q. Quiroga, C. Koch, and I. Fried. On-line, voluntary control of human temporal lobe neurons. *Nature*, 467:1104–1108, 2010.

[2] T. Cukur, S. Nishimoto, A G Huth, and J. L. Gallant. Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16:763–770, 2013.

[3] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. *Rethinking Innateness: A connectionist perspective on development*. MIT Press, Cambridge, Massachusetts, 1997.

[4] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.

[5] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[6] S. Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[7] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 3rd edition, 2001.

[8] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1106–1114, 2012.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of IEEE*, 86(11):2278–2324, 1998.

[10] J. L. McClelland, D. E. Rumelhart, and The PDP Research Group, editors. *Parallel Distributed Processing*, volume 2. MIT Press, Cambridge, Massachusetts, 1986.

[11] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh. *Computational Maps in the Visual Cortex*. Springer, Berlin, 2005.

[12] M. Rigotti, O. Barak, M. R. Warden, X. Wang, N. D. Daw, E. K. Miller, and S. Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497:2003, 2003.

[13] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group. *Parallel Distributed Processing*, volume 1. MIT Press, Cambridge, Massachusetts, 1986.

[14] L. K. Samuelson and L. B. Smith. They call it like they see it: spontaneous naming and attention to shape. *Developmental Science*, 8(2):182–198, 2005.

[15] J. Sharma, A. Angelucci, and M. Sur. Induction of visual orientation modules in auditory cortex. *Nature*, 404:841–847, 2000.

[16] P. Voss. Sensitive and critical periods in visual sensory deprivation. *Frontiers in Psychology*, 4:664, 2013. doi: 10.3389/fpsyg.2013.00664.

[17] Y. Wang, X. Wu, and J. Weng. Synapse maintenance in the where-what network. In *Proc. Int'l Joint Conference on Neural Networks*, pages 2823–2829, San Jose, CA, July 31 - August 5, 2011.

[18] J. Weng. Why have we passed "neural networks do not abstract well"? *Natural Intelligence: the INNS Magazine*, 1(1):13–22, 2011.

[19] J. Weng. *Natural and Artificial Intelligence: Introduction to Computational Brain-Mind*. BMI Press, Okemos, Michigan, 2012.

[20] J. Weng. Brain as an emergent finite automaton: A theory and three theorems. *International Journal of Intelligent Science*, 5(2):112–131, 2015.

[21] J. Weng, N. Ahuja, and T. S. Huang. Cresceptron: a self-organizing neural network which grows adaptively. In *Proc. Int'l Joint Conference on Neural Networks*, volume 1, pages 576–581, Baltimore, Maryland, June 1992.

[22] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation using the Cresceptron. *International Journal of Computer Vision*, 25(2):109–143, Nov. 1997.

[23] J. Weng and M. Luciw. Dually optimal neuronal layers: Lobe component analysis. *IEEE Trans. Autonomous Mental Development*, 1(1):68–85, 2009.

[24] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291(5504):599–600, 2001.