Telecom Business Review
12 (1) 2019, 18-22
http://publishingindia.com/tbr/

Layered Approximation for Deep Neural Networks

Utpal Chakraborty

Chief Data Scientist, Head of Artificial Intelligence, Yes Bank, Mumbai, Maharashtra, India.
Email: utpal.chakraborty@yesbank.in

Abstract

Artificial Intelligence has created immense hype in the last decade and the credit for the same goes to the groundbreaking breakthrough named "Deep Learning" or "Deep Neural Network". Although "Artificial Neural Network", the foundation of Deep Learning as a concept has been prevalent since 1958 but the actual implementation for solving real business use cases have only been possible over the last decade. Deep Neural Networks has demonstrated significant results in the fields of computer vision, speech recognition, and machine translation, and outperformed human brain in many instances. Artificial Neural Network, as it is inspired from human biological neural superstructure has few structural similarities but not possible in terms of how the human brain or biological neural network works because we have still limited information on its functioning. Nevertheless, Deep Neural Network paved the way for many possibilities and currently it is the most promising technology that we have in the field of Artificial Intelligence.

Keywords: Deep Neural Network, Artificial Intelligence, Artificial Neural Network, Layered Approximation

1. Fundamental Limitations of Artificial Neural Network

Although Artificial Neural Network has numerous advantages but they also have a couple of fundamental limitations. These limitations can be broadly classified into two categories-

- *Scientific or Mathematical Limitations and*
- *General Limitations*

Scientific or mathematical limitations include issues with- "Piece-wise Linear Curve", "Flat Activations", "End-To-End-Learning", "Learning Many Orthogonal Functions" and actual validation of "Universal Approximation Theorem" with a single layer Perceptron which are beyond the scope of this research paper. But there are other general limitations as well as -

- Requirement of a huge relevant dataset for training the deep neural networks (Jason Browniee, 2019).
- Requirement of huge computational power for training and substantial time requirement for training such models.
- Hyper-parameter tuning for the deep learning models is sometimes a tedious process although some automation of such activities are now available but still needs maturity (we know very little about human brain, n.d.).

Few of the comments from experts on the limitations of deep neural network are:

- Francois Chollet, Google, inventor of Keras library said - *"Current supervised perception and reinforcement learning algorithms require lots of data, are terrible at planning, and are only doing straightforward pattern recognition"*.
- Similarly, 'Intriguing Properties of Neural Networks' by Christian Szegedy from Google and others commented like - *"We can cause the network to misclassify an image by applying a certain noise, which is found by maximizing the network's prediction error"*.
- Another comment from the expert community is - 'Deep Networks are Easily Fooled' by Anh Nguyen at the University of Wyoming - *"Changing an image of a Lion in a way imperceptible to humans can cause a Deep Neural Network to label the image as something else entirely mislabeling, a lion a Library"*.

Also, "Universal Approximation Theorem" which claims that a neural network with a single hidden layer containing a finite number of neurons can approximate any continuous function (Huang, Chen, Siew, 2006).

This implies - for any problem, there exists a neural network which can approximate it with some degree of

accuracy. But unfortunately, it provides little information about which classes of functions can they approximate well. Why some of them work unreasonably well and some others not at all. And what depth and size of a network are sufficient to guarantee an approximation in a particular scenario?

Moreover, it is unclear how to design a very deep neural network effectively. Many times arbitrarily adding more layers does not help rather worsen the performance of the network which has been observed in many practical scenarios.

But more than all the above, Artificial Neural Network and for that matter, the AI itself is lagging the most important and the most aspired capability; what is called the "General Intelligence" which is far beyond what we have achieved with AI till now, the "Narrow Intelligence".

To overcome the above mentioned limitations and the limitations imposed by "Back Propagation" process namely - "Vanishing Gradient" and "Exploding Gradient" in many cases where the network essentially ceases to learn further because the gradient value either becomes too negligible or too big; researchers has also suggested few alternatives such as *"Geoffrey Hinton's - Neuron Capsules", "Yann LeCun's - Energy-based models" and "Zhi-Hua Zhou's - Deep gcForest"* but none of these has been turn out to be that effective neither adopted well till now.

An AI system can be far better than human in performing some activity in the single domain it is made for but it doesn't have the awareness of what it is performing. Also it can't bring in the experience from some other context (cross-domain experience) to the current context or relate experiences. Human brain can do this very well and many more amazing things which are normally referred as "General Intelligence".

An effort has been made in this research paper to refer back to the very fundamental elements and principles on which human neural system works and reference those back in the case of Artificial Neural Network and try to address some of the fundamental limitations.

Recent research in neuroscience has revealed the fatigability of neurons, meaning neurons has a tendency to get tired or lose strength very quickly. Neurons are electrically polarized and exhibit an electric potential difference of only "-70" Micro Volt with their surroundings and they are capable of sending weak electric signals. Due to which they can't be the center of human cognition, they can probably be the smallest units in the human cognition system. Rather a cortical map is a set of thousands of cortical columns and each column a set of about 100,000 neurons claimed to be the center of cognition for human brain. This is again on a very higher abstract level than the extremely granular neurons level. So cortical columns should be considered as the building blocks of cognition and intelligence.

If we consider this to be factual and apply the similar approach in Artificial Neural Network as well, will apparently solve some of the fundamental limitations, like requirement of huge training data, huge computational power and significant training time. In this case the layers of artificial neurons will be abstracted to be the center for artificial cognition rather than the individual neurons. The group of neuron layers in a layer can be visualized as network-of-networks.

2. Biological Neural Network

The nervous system is the central part of the human body which transmits signals back and forth to different parts of the body that helps coordination of our voluntary and involuntary movements. At a cellular level the nervous system consists of a special type of cell called "neuron". There are near about 100 billion neurons in human brain and they are connected to each other through a wire like structure called Synapse, which acts like a pathway for electrical signals to travel from one neuron to the other.

Arguably, neurons - the tiny polarized cells inside the brain are considered to be the "Center of Cognition" of human.

As per the information available, the protein synthesis occurs at the nucleus of the biological neuron called "Soma" considered to be the process responsible for human cognition. The signals emitted by the neurons from one to the other through Synapse are believed to be partly electrical and partly chemical that is the byproduct of the phenomena that happens due to the protein synthesis inside it.

Human neurons receive the inputs to be processed from the outside world through our five senses - ear, eye, nose, skin and tongue. The inputs received through these organs are encoded into electrical signals before being sent to the layers of neurons to be processed. The biological neurons works on a "Threshold Based" principle. This means whenever the input signal received by a neuron after its chemical synthesis crosses a particular electrical

threshold limit, it fires the signal to the next neuron. The next neuron then takes this signal as its input and does the same process which continues through millions of subsequent layers of neurons until arrives at a conclusion on the piece of information that has been processed.

3. Comparison - Biological vs. Artificial Neural Network

The Artificial Neural Network works very similarly in the same fashion, "Threshold Based" principle called activation functions. The difference primarily lies between how both these networks learn new things or the learning process.

Scientists have revealed that child's brain at a very early stage from zero to one year used to have around 100 billion neurons and only few million connections. But these connections exponentially multiplies during the development stage of the brain till the age of 6 (six) years at the rate of around 1000 (thousand) connections per second. That's how a child learns and the learning process is essentially through building more and more connections among the neurons. An interesting fact about the human brain is its constantly changing, constantly evolving and establishing millions and billions of new connections, redefining existing connections and removing irrelevant connections among the neurons across different parts of our brain. This is the process how human brain learns new things, adds experience to those learnings and draws conclusion on the incidents about our life.

In case of Artificial Neural Network the learning process is called "Back Propagation", is a process of optimizing the connections among the artificial neurons from backward by adjusting the "weight" and "biases" variables and finally move as close to the ideal state and the golden matrix of weights and biases where the network is optimized to provide the best possible predictions.

4. Layered Approximation - Philosophy

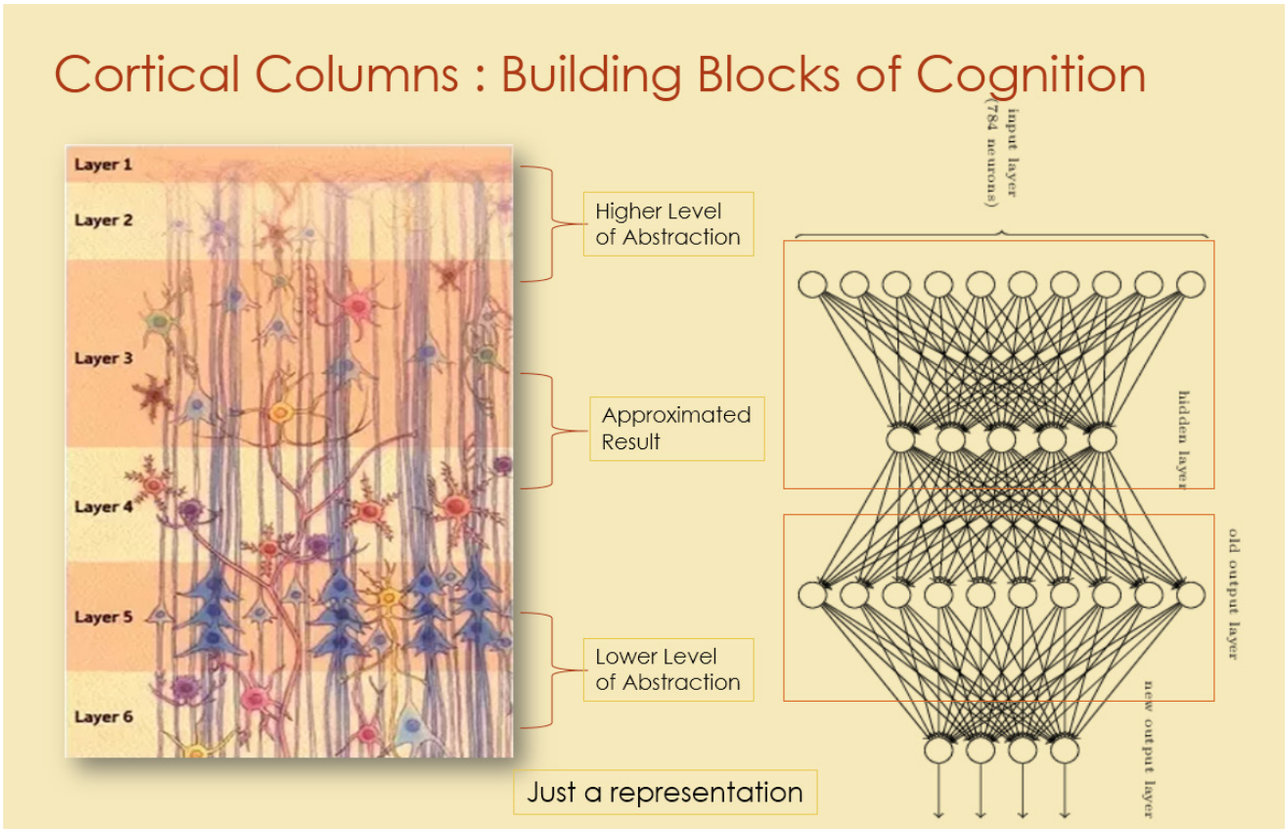

Fig. 1: Cartical Columns - Building Blocks of Cognition

"Layered Approximation" for Deep Neural Network is a design principle inspired from the real world fact that "Everything we see, we do and we perceive in this universe are outcomes of approximated functions, approximated actions and approximated decisions and none of those are absolute in nature".

Aristotle once wrote - "It is the mark of an instructed mind to rest satisfied with that degree of precision which the nature of the subject limits, and not to seek exactness where only an Approximation of the truth is possible".

Layered Approximation principle also advocates that whatever operations (chemical or mathematical) that happens inside human brain are approximated operations and generates approximated results in a layered (zoom-in / zoom-out) fashion. If the complexity of the problem in hand increases, the operations to deal with such more complex problems leads to more and more minute or detailed operations which demands zoom-in to the situation hence more computation and as more detailing involved. In case the problem in hand can be dealt with higher level of abstraction (zoom-out), the operations take place inside the brain for such scenarios are on higher abstract level and requires less computation and detailing.

A classic example is, we can easily drive a car in an empty road for couple of hours at a stress without having much mental exhaustion because the operations happen in your brain (inside the neuron layers) in this scenario are on a much higher level of abstraction which requires less computation and processing. Whereas if you have to drive in nasty traffic, although you will probably travel much lesser distance and spend lesser time but you will be tired and stressed out even driving lesser. This is because, here in this scenario your brain has to take into consideration every meter, every minimum curve while driving in such narrow space and hence the operations happen inside the brain with lesser abstraction and with more detailing.

Another good example, we can do work which requires higher abstraction and less computation for tasks like writing a poem or writing a novel. Probably we can keep writing at a stress for couple of hours. But when it comes to solving a complex mathematical problem, perhaps it will be very difficult to concentrate constantly for an hour or so and even then we get fully stressed out.

Through abstraction and approximation human brain optimizes its computational resources for many complex operations which otherwise would have not been possible considering its tiny size.

Similarly while learning new things human brain takes an abstract approach. Unlike Artificial Neural Network which learns pixel-by-pixel from an image; human brain creates an abstract image of the actual image for learning. Likewise for recognizing an image, human brain tries to map it with abstraction first to quickly recognize it fast and probably then goes into details by zooming in if the situation demands detailing. So essentially there is a "Decision Making System (DMS)" that works together in collaboration to decide what level of abstraction is necessary for the given situation and where to stop and where more dip down is required.

Augmenting "Decision Making System (DMS)" along with the Deep Artificial Neural Network is core of Layered Approximation approach. The DMS is going to aid in deciding at which particular layer the approximation performed by the network is sufficient and acceptable for that particular problem in hand. This approach is going to reduce the requirement of huge computational capability, huge training data and significant training time to train the network drastically.

A real life example is it's because of approximation our brain can be of that sure identifying a person or any object even looked at from a distant acute angle, low light for just a fraction of a second. Even with that very minimum information "Decision Making System (DMS)" can essentially be sure of to take a decision that you have seen that person or the object. Of course that minimum information should be vital information to distinguish and identify that particular person or the object with the degree to certainty expected by DMS. Without the involvement of "Decision Making System (DMS)" it will not be possible for the brain to provide a prediction with that little information in those scenarios.

5. Practical Approach for Implementation

Designing and Testing a Deep Neural Network for any specific requirement is always a challenge and does not guarantee that it will always provide the same level of accuracy for different data sets. As the deep network grows in size and depth the testing complexity becomes almost unmanageable. Also, the hyper parameter tuning becomes more and more complex. All these challenges come from limitations of the very fundamental principles on which Neural Network training process is based on, the Back Propagation.

Proposed Layered Approximation is a design principle which can be applied to a deep neural network during its design that can address the training & testing challenges to a great extent although it can't solve all the fundamental challenges at least at its present stage.

With Layered Approximation design approach applied to a Deep Neural Network, the requirement of huge training data and huge computational power can also be addressed to a certain extent and thus the reduction in training time and testing complexities.

References

Huang, G. B., Chen, L., & Siew, C.-K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks, 17*(4), 879-892. Retrieved from http://www.ntu.edu.sg/home/egbhuang/pdf/I-ELM.pdf

Browniee, J. (2019, January 2). Impact of dataset size on deep learning model skill and performance estimates. better deep learning. Retrieved from https://machinelearningmastery.com/impact-of-dataset-size-on-deep-learning-model-skill-and-performance-estimates/

LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., & Huang, F. J. (2006). A tutorial on energy-based learning. In G. Bakir et al., *Predicting Structured Data.* MIT Press. Retrieved from http://yann.lecun.com/exdb/publis/

Nielsen, M. A. (2015). Neural networks and deep learning. Determination Press. Retrieved from http://neuralnetworksanddeeplearning.com/chap4.html

Sara, S., Nicholas, F., & Hinton, G. E. (2017). *Dynamic routing between capsules.* 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. Retrieved from https://arxiv.org/pdf/1710.09829.pdf

We Know Very Little About Human Brain. (2018, August 19). Retrieved from https://mc.ai/we-know-very-little-about-human-brain-artificial-neural-network-vs/

Zhou, Z. H., & Feng, J. (2017). *Deep forest: Towards an alternative to deep neural networks.* Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Retrieved from https://arxiv.org/abs/1702.08835