01 Jan 2018

Outsmart Moore's Law with Machine Learning

Dustin Tanksley

Donald C. Wunsch
Missouri University of Science and Technology, dwunsch@mst.edu

Recommended Citation

D. Tanksley and D. C. Wunsch, "Outsmart Moore's Law with Machine Learning," *IEEE Eta Kappa Nu The Bridge Magazine*, vol. 115, no. 2, pp. 6-9, Institute of Electrical and Electronics Engineers (IEEE), Jan 2018.

Outsmart Moore's Law with Machine Learning

by: Dustin Tanksley and Donald C. Wunsch

Over the last half century, computing has transformed most aspects of society due to a rapid increase in computation power. With the possible end of Moore's Law in sight, much of this growth could come to an end. This paper will discuss why machine learning will continue growing even after Moore's Law, and demonstrate why it is a great time to enter the field.

WHAT IS MACHINE LEARNING?

At the most fundamental level, machine learning is the process of using advanced function approximators and large amounts of data to create a mathematical representation of a problem. As an example, one could take a group of pictures of cats and dogs, and identify them. To do this, some function would have to map these pictures to a numeric value, possibly 0 for cats and 1 for dogs. The art of machine learning is creating these complicated functions and then apply meaning to the mathematical representations. Three major approaches to such tasks are supervised learning, reinforcement learning, and unsupervised learning.

Supervised learning is typically the most intuitive. In this type of learning, labeled data are fit to an appropriate function; for example, matching the price of a house to the size, other factors. For simple problems, a linear (or higher order) regression algorithm works just fine, however as more parameters such as; bedroom and bathroom count, location within or distance to a major city, population density, and other factors are considered, nonlinear models are often needed. Nonlinear versions of regression exist, but neural networks and other methods are often competitive with those, and the whole family of such approaches can be considered types of machine learning. Perhaps the best example of supervised learning success can be seen in ImageNet, an effort to identify the object in an image. ImageNet consists of more than 10 million internet images that have been identified and labeled by humans, and as of 2017, the best algorithm achieved a 97.7% correct classification, which is better than most humans (typically 90-95%) [1][2].

Reinforcement learning typically uses similar neural network architectures as supervised learning, with the key difference that data usually must be generated/gathered, so that performance functions replace the role of labels. Reinforcement learning is easiest to visualize in games, such as tic-tac-toe. Several actions are available, allowing the agent to place a mark in one of the squares, and in doing so generates a new data-point, but the agent does not know if this was a good or bad move. When the game ends, the agent is told if it wins, loses, or draws and must use this data to determine if all the actions it took were good or bad. Classifying moves as good or bad is a somewhat difficult process, but many improvements have been made in the field, with the most recent being AlphaGo, an effort by Google to master the game of Go (a far greater computational challenge than Chess). AlphaGo's most recent achievements include beating a former world champion in a 5-game match (AlphaGo Lee) and beating 60 of the top Go players in the world without any losses (AlphaGo Master). Beyond this, an even stronger version has been released, AlphaGo Zero, which defeated AlphaGo Master 89-11 and is notable for being completely trained by reinforcement learning from playing, without the benefit of any initial supervised learning [3][4].

Unsupervised Learning is very different from supervised and reinforcement learning, in that it does not have a training target. Clustering, the most common form of unsupervised learning, groups inputs together based on similarities, and determines it has succeeded based on how tightly the groups are packed, and how many outliers are present. In this way, unsupervised learning is typically very good at finding patterns in the data, though with complex datasets, the pattern found may be difficult to interpret. While this can make it unsuitable for some of the problems in supervised and reinforcement learning, it does have some very profound applications. For example, it can be used to divide and conquer other problems such as the combinatorically-demanding Traveling Salesman Problem. Unsupervised learning was used as a heuristic to divide the problem and achieve a dramatic speedup over the best previous solution. [5] Whether the final processing is done by another algorithm or a human, reducing the complexity of data analysis is among many applications of unsupervised learning.

Overall, each method to machine learning has its own strengths and weaknesses. It is clear that humans exhibit traits from each, being able to learn from either a teacher (supervised), or trial and error (reinforcement), or being able to classify objects based on their similarities to find patterns (unsupervised). While human level AI may be far away, these methods can be applied to most problems with very good results.

WHY WILL MACHINE LEARNING OUTLIVE MOORE'S LAW?

While Moore's law has certainly helped machine learning, it is not needed for the continued growth of the field. Algorithmic advances are continually improving the field, with new methods for learning along with better parallel processing constituting the most significant increase in performance. Furthermore, increasingly specialized hardware that focuses on the operations needed for machine learning has led to tremendous growth.

To put this in perspective, IBM Deep Blue managed to defeat the reigning world champion of chess, Garry Kasporov, in 1997. Deepmind's AlphaGo Lee defeated Lee Sedol, 18-time world champion of Go, in 2016. Chess has a state space complexity of 10^47, while Go has a state space complexity of about 10^170. This means in those 19 years, computational efficiency would have increased by 123 orders of magnitude, or 297,600,000% per year. While this figure is very approximate, it highlights just how impressive the growth of machine learning is.

To further highlight the growth of AlphaGo particularly, the original version used 176 GPUs, and required months of training. When it defeated Lee Sedol, it had switched to 48 Tensor Processing Units (TPUs), which are optimized for machine learning. Just months later Deepmind launched AlphaGo Zero, which started learning without any human game data, and in 3 days was stronger than the version that beat Lee Sedol (AlphaGo Lee), and ultimately after 40 days was fully trained, far outmatching any previous results. This newest version only runs on 4 TPUs, and even despite this is many times stronger then AlphaGo Lee.

Just as algorithmic advances have accelerated machine learning, newer, more optimized hardware has made its impact as well. When self-driving car research took off around 2010, GPU acceleration had started to become mainstream. Since then, programs that could utilize these resources became more common, and has since become the de-facto standard for machine learning. The more specialized TPU that is aimed solely at deep learning applications has finally stepped into consumer grade products with Nvidia's release of the Titan V, which claims up to 110 TFlops of compute power in deep learning applications. This shows a large step towards creating specialized hardware specifically for machine learning, and the potential demand for such systems.

THE BARRIERS TO ENTRY ARE COMING DOWN.

Getting into machine learning has become much easier over the past decade. Many APIs are available for all programming languages. With a host to choose from including; Tensorflow, Caffe, Keras, DeepLearning4J, or even MATLAB's Machine Learning Toolbox, it requires very little time and fairly little code to start applying machine learning. Many tutorials are available to help people get started with machine learning, and while there will be some learning time for people unfamiliar with the subject, very little knowledge of machine learning algorithms is needed to successfully use these tools.

Another barrier that has been coming down is the need for supercomputing. While the CPU in a standard computer can usually run most algorithms especially after they are trained, to effectively train on large amounts of data in a reasonable time requires much more power. Fortunately, many APIs nowadays are automatically GPU accelerated, allowing for massive speedup. Specialized hardware is even reaching consumer levels, as can be seen in products such as the Intel Movidius, a USB stick that can be used to accelerate machine learning. As well, several cloud services from companies such as Amazon and Google offer GPU and TPU nodes for use, allowing for models to be trained without buying hardware.

It is important to note that these growths are not a direct result of Moore's Law. While Moore's law clearly benefits newer hardware, allowing for more transistors in the same area, GPUs don't necessarily rely on these and many of their gains are simply from increasing core count. Furthermore, specialized hardware simply boils down to configurations of the transistors that allow for faster computation, at a cost in speed for general purpose computation.

CONCLUSION

The world is fast becoming AI centric as autonomous vehicles are becoming more of a reality every day, autonomous drones/delivery bots are being developed, robots are becoming more adaptable, and the ability of computers to sell to us is growing dramatically. With the rapid advancement of machine learning into nearly every area, it is fast becoming commonplace. With the continued innovations in machine learning algorithms and specialized hardware, the growth in this field will continue far beyond the limits of Moore's Law.

REFERENCES

[1] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int J Comput Vis, vol. 115, no. 3, pp. 211–252, Dec. 2015

[2] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," arXiv preprint arXiv:1709.01507, 2017.

[3] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

[4] D. Silver et al., "Mastering the game of go without human knowledge," Nature, vol. 550, no. 7676, p. 354, 2017.

[5] S. A. Mulder and D. C. Wunsch, "Million city traveling salesman problem solution by divide and conquer clustering with adaptive resonance neural networks," Neural Networks, vol. 16, no. 5–6, pp. 827–832, Jun. 2003.

Donald Wunsch is the Mary K. Finley Missouri Distinguished Professor at Missouri University of Science and Technology (Missouri S&T).

He is the Director of the Applied Computational Intelligence Laboratory, a highly multidisciplinary research group. Earlier employers were: Texas Tech University, Boeing, Rockwell International, and International Laser Systems.

His education includes: Executive MBA - Washington University in St. Louis, Ph.D., Electrical Engineering - University of Washington (Seattle), M.S., Applied Mathematics (same institution), B.S., Applied Mathematics - University of New Mexico, and Jesuit Core Honors Program, Seattle University.

Key research contributions are: Clustering / Unsupervised Learning; Biclustering; Adaptive Resonance and Reinforcement Learning architectures, hardware and applications; Neurofuzzy regression; Traveling Salesman Problem heuristics; Games; Robotic Swarms; and Bioinformatics.

He is an IEEE Fellow, previous INNS President, INNS Fellow, NSF CAREER Awardee, 2015 INNS Gabor Award recipient, and Eta Kappa Nu member. He served as IJCNN General Chair, and on several Boards, including the St. Patrick's School Board, IEEE Neural Networks Council, International Neural Networks Society, and the University of Missouri Bioinformatics Consortium, Chaired the Missouri S&T Information Technology and Computing Committee as well as the Student Design and Experiential Learning Center Board.

He has produced 20 Ph.D. recipients in Computer Engineering, Electrical Engineering, Systems Engineering and Computer Science; has attracted over $10 million in sponsored research; and has over 450 publications including nine books. His research has been cited over 15,000 times.

Dustin Tanksley is a PhD Engineering at Missouri University of Science and Technology.

He is a member of the Applied Computational Intelligence Laboratory, focusing his research on game theory, reinforcement learning, and neural networks.

He has previously earned an associate's degree in Science from the Missouri Academy, and a bachelor's degree in Computer Engineering from Missouri University of Science and Technology, as well as being the youngest person to pass the Qualifying Exam in Computer Engineering at Missouri S&T.

He has participated in several chess AI competitions and has placed several times in the top three. He has been recognized for his achievements on several occasions, notable receiving both a GAANN Fellowship and Chancellor's Distinguished Fellowship. He is an active Gamma Theta member of HKN.