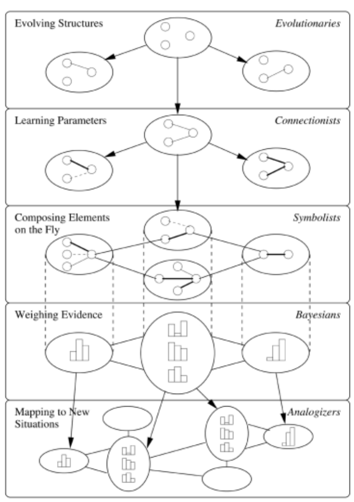
Paper 1
Theo Shin

The Master Algorithm: Understanding the Complexity of Machine Learning

Upon reading Pedro Domingo's *The Master Algorithm* and Max Tegmark's *Life 3.0*, machine learning and artificial intelligence were reminiscent of "pie in the sky" concepts requiring nothing short of a miracle to bridge the gap between far-fetched ambitions and an actual reality where automation rules. Recent surges in technology have demonstrated remarkable improvements in gadgets and applications especially in smart phones, voice and facial recognition, self-driving cars, and virtual reality. Is it realistic to imagine a world where machines could replace aircraft pilots, nannies to watch your children, and Terminator machines to replace the military's front line, potentially sparking a judgement day. Pragmatically speaking, I do see the practicality in employing machines to man hazardous type jobs, such as plant workers, to eliminate risk of injury or death. In Domingo's *The Master Algorithm,* the author believes how soon these advancements can become fully realized depends on solving the "master algorithm," which he suggests needs to be derived from the perspectives of Symbolists, Connectionists, Evolutionaries, Bayesians (in reference to Naïve Bayes expressions), and Analogizers (Refer to Figure 1):

Figure 1: "Structure for a Master Algorithm with Five Capabilities" (Reference: Domingo p. 33, 2015)



From these five perspectives, the master algorithm must incorporate learning mechanisms through natural selection, uncertain inference, a brain-like processor (through reverse engineering), and recognition of similar experiences. The approach to machine learning then becomes what can ultimately be made or done better through a learning algorithm and is the world better off because of it. Not surprisingly, understanding the complexity of machine learning and its purpose comes down to looking at the end state and deriving an algorithm able to process vast amount of incoming information and discern which data is important. Although the thought of deriving such an algorithm appears overly complex, according to Domingo, machine learning has a delightful history of simple algorithms unexpectedly beating intricate ones. Domingo compares the eventual master algorithm to the significance of what the “hand is to pens, swords, screwdrivers, and forks” (Domingo, 40). Of course, this master algorithm will be heavily influenced by Alan Turing’s machine since “every conceivable problem that can be solved by logical deduction was solved by a Turing machine. The Master Algorithm is for induction, the process of learning, what the Turing machine is for deduction. It can learn to simulate any other algorithm by reading examples of its input-output behavior” (Domingo, 33).

So, the objective of a master algorithm is constructing a program that isn’t overly complex and yet, can be applied to solve the world’s host of problems. Starting with Bayesians, Domingo states “Naïve Bayes,” expressed as a short equation, could be implemented in everything ranging from diagnosing patient conditions to learning spam filters, two implementations seemingly having nothing to do with each other and all in a fraction of the time it would take experts to figure out each of those scenarios. In statistics, Bayes’ theorem is a single formula conceptually underlying all learning. How it can be applied to machine learning starts with how a Bayesian learner can take a set of hypotheses and continually update/remove from the set depending on what it determines to be compatible or not, thus resulting in a single hypothesis (Domingo, 23). According to Domingo, Bayes’ theorem is a machine that turns data into knowledge and according to Bayesian statisticians, it’s the only correct way to turn data into knowledge.

The Impact/ Role of Data

Now, let’s take a look at data. The semi-recent “big data” revolution points to the enormous importance of data and how data wrangling can often change and lead to better decision making. So much can be done with data and with the enormous influx of data continuing to grow, the potential is enormous for the application of data science. In Seth Stephens-Davidowitz’s *Every Body Lies*, the author’s talking points include how data significantly strengthens research capabilities. For example, when using data to compare different geographical regions, multiple observations in a dataset can now be zoomed in with clarity similar to how a photo needs a lot of pixels in order to be able to zoom in on a smaller portion of it (Stephens-Davidowitz, Ch. 5). Domingo adds human intuition can’t replace data (He references Michael Lewis’s “Moneyball,” where statistical analysis beat talent scouts in baseball; Domingo, 39). Working with big data allows analysts in numerous sectors to obtain and use the information in further possible ways. Consider this example from *Every Body Lies*, where the question is what are the chances a person born with parents in the bottom 20 percent of the income distribution reaches the top 20 percent of the income distribution as an adult within the United States compared to other countries. The traditional way to answer this question is to look at a representative sample of Americans and compare this to similar data from other countries. With the capabilities of working with so much data, data analysts are now able to zoom in on smaller demographics of people, moving from city to city, comparing the statistics of every U.S. area and its effects, and allows for testing causation and not just correlation. Stephens-Davidowitz states the analysis concluded the U.S. is very much a collection of societies, some of which are ‘lands of opportunity’ with high rates of mobility across generations, and others in which few children escape poverty (Davidowitz, Ch. 5).

Throughout our course, it became clear not only how essential data is but how data can compound the issue. For instance, in our exercises for doubling love each day and calculating the possible amount of chess moves, we observed how quickly the upper bound increases, along with the computational resources and time to derive a solution. Domingo touches upon similar notions detailing how combinatorial explosions can bring fear to many computer scientists and states machine learning is a race between the amount of data and the number of hypotheses considered. Which methods can be utilized to combat combinatorial explosions? Possibly with the Probability Approximately Correct (PAC) analysis. According to a Wikipedia article, Leslie Valiant came up with PAC as a framework where after a learner receives samples it must select a hypothesis which will result in low generalization error and with high probability. The learner is expected to find efficient solutions given time and space constraints and implement an effective procedure within those bounds using mathematical analysis. When we go back to Bayesian methods for constructing a solution, Domingo explains while inverse deduction has its limitations in being computationally intensive (Increasing the difficulty with scaling massive data sets), Symbolist's choice of algorithm is "decision tree" induction where rules are created to match an instance (Domingo, 85).

In part, working with big data sets appears to be somewhat of a paradigm when being implemented with a universal learner. On the one hand, data is absolutely necessary to constructing an algorithm which can help "cure" our woes and enhance human existence. On the other hand, the complexity in working with enormous data sets can be overly burdensome. Domingo writes, "from a million runs of a thousand coin flips, it's practically certain that at least one run will come up all heads, and a million is a fairly small number of hypotheses to consider" (Domingo, 74). This familiar example was demonstrated in our second problem set when learning about the "regression-to-the-mean" expressed in Kahnemann and Tversky's theory. Domingo captures some of this complexity in his story from Jorge Luis Borges titled, "Funes the Memorious," when he talks about a youth with perfect memory. Domingo states some might see this as an enormous advantage, but the story goes on to show how it is a curse. Funes, the youth, can remember the exact shape of clouds in the sky at an arbitrary time, but has undeniable trouble in understanding that a dog seen from the side at 3:14 p.m. is the same dog seen from the front at 3:15 p.m (Not at all dissimilar from how robots cannot accurately discern from optical illusions). For Funes, two things are only the same if they look the same down to every last detail and tying this to the master algorithm, an unrestricted rule learner is unable to function without constraints.

Use of Data for Machine Learning / Application of Data in Understanding Machine Learning

So, what can we make from this data and how can a machine "learn." According to Max Tegmark in *Life 3.0*, the human brain's wetware is akin to hardware and the software would need to include human experiences and the knowledge obtained from processed information derived from senses. Everything from the ability to recognize friends to abilities learned (walking, writing, singing, telling jokes, learning a language) were not built into our brains during birth. Children have their childhood curriculum largely designed by family and school, but as the child grows older, he gains more power in deciding what to learn and thus formatting the software (Tegmark, Ch. 1). To get an idea of how experiences can be expressed in the computing world, Domingo states the S curve is really like the jack-of-all-trades in mathematics and variances in its phase transitions are akin to a human's mood swings in events such as birth, adolescence, falling in love, being married, moving to a new town, retiring, and eventually dying (Tegmark, Ch. 1). Mathematically and computationally, human “experiences” can be written and expressed in a program. How then can machines continue to learn with pre-installed software (I am reminded of the movie, “Ex Machina,” a modern take on the Turing test). According to *Life 3.0*, the human synapses store all of the knowledge and skills as roughly 100 terabytes worth of information, while the DNA stores merely a gigabyte (Tegmark, Ch. 1). Whether this is indeed accurate, designing a learning computer to mimic human brain will invariably involve the master algorithm.

In *The Master Algorithm*, Domingo discusses how a team of neuroscientists from MIT were able to rewire the brain of a ferret, reroute its connections from its eyes to the auditory cortex and successfully reroute the connections from the ears to the visual cortex. The conclusion drawn is the brain uses the same learning algorithm throughout, with different senses distinguished only by the different inputs they are connected to. As mentioned in the previous paragraph, the master algorithm will likely need to incorporate from the human brain as it is responsible for everything imagined and perceived and necessary for acquiring new experiences and skills. Designing an algorithm based on the human brain will most definitely have its own host of complex hurdles. For one, most of the brain is devoted to sensing and moving, and in order to build a brain-based algorithm, we must borrow aspects of it evolved for language (Domingo, 31). As discovered in our course, teaching a machine or writing an algorithm to learn what humans innately do or perceive as easy can be particularly challenging for machines.

In directly comparing human hardware to computers, Domingo brings up Hebb’s rule (In reference to Donald Hebb, author of *The Organization of Behavior,* who coined the phrase “Neurons that fire together wire together.” Connectionist’s viewpoint is that all neurons learn simultaneously and mirror the different properties of computers and brains. Computers are extremely fast, switching transistors on and off a billion times per second, but are limited to doing everything one at a time. In contrast, the brain performs many computations in parallel with billions of neurons working at the same time, each of the computations is comparatively slow since “neurons can fire at best a thousand times per second” (Domingo, 94). Domingo continues:

“The number of transistors in a computer is catching up with the number of neurons in a human brain, but the brain wins hands down in the number of connections. In a microprocessor, a typical transistor is directly connected to only a few others, and the planar semiconductor technology used severely limits how much better a computer can do. In contrast, a neuron has thousands of synapses. If you’re walking down the street and come across an acquaintance, it takes you only about a tenth of a second to recognize her. At neuron switching speeds, this is barely enough time for a hundred processing steps, but in those hundred steps your brain manages to scan your entire memory, find the best match, and adapt it to the new context (different clothes, different lighting, and so on). In a brain, each processing step can be very complex and involve a lot of information, consonant with a distributed representation” (Domingo, 94). Even still, as complex as the brain is, a universal master algorithm does not need to represent a realistic model of the brain and simply should be kept efficient as possible.

Application of Evolutionary / Innate Learning

Lastly, we will discuss evolution as an algorithm. Domingo discusses how human DNA has been running and modified for over three billion years on the most powerful computer, Earth. Machine learning's version of nature versus nurture comes down to which model is better suited for the master algorithm, evolution or the brain. As nature and nurture combine to produce humans, perhaps the algorithm must contain elements of both (Domingo, 28). For a computer, nature is the programming it runs while nurture is the relevant data acquired. Domingo states having something like a "fitness" function is a no-brainer. Needing a program to correctly diagnose a patient, "one that correctly diagnoses 60 percent of the patients in our database is better than one that only gets it right 55 percent of the time, and thus a possible fitness function is the fraction of correctly diagnosed cases" (Domingo 123). Luckily, according to Domingo, genetic algorithms implicitly contain strings with an exponential number of building blocks (schemas), which results in a more efficient search. In this case, the combinatorial explosion works in favor rather against. The key to running an efficient universal learner must entail a combination of an established brain and its ability to encode the nurturing aspect and fill the brain with vast amounts of information. Thus, it would need to take into account structured learning and establishing weights for which information is most important, similar to the Bayesian model. Like "survival of the fittest," the learning algorithm will invariably pick and choose between which "neural connections" are best suited as connections continue to grow stronger and other paths diminish.

The applicability for machine learning are vast and none more important than cancer research. According to Domingo, a master algorithm may be able to help discover the cure. Domingo states a vast amount of patient and drug data combined with knowledge mined from biomedical literature is the pathway to a cure and more generally, inverse deduction will be the method, similar to how new knowledge is gleaned in biology (Domingo, 42). Domingo states, "learning which drugs work against which mutations requires a database of patients, their cancers' genomes, the drugs tried, and the outcomes. The simplest rules encode one-to-one correspondences between genes and drugs, such as If the BCR-ABL gene is present, then use Gleevec (BCR-ABL causes a type of leukemia, and Gleevec cures it in nine out of ten patients.) Once sequencing cancer genomes and collating treatment outcomes becomes standard practice, many more rules will be discovered" (Domingo, 83).

Conclusions

Does a master algorithm no matter how great the achievement ultimately make the world a better place? Domingo mentions a prominent machine-learning skeptic, a linguist by the name of Noam Chomsky. Chomsky believed that language must be innate, because the examples of grammatical sentences children hear are not enough to learn a grammar. This only puts the burden of learning language on evolution. However, it does not argue against the master algorithm but only against it being something like the brain. Chomsky seems to equate machine learning with behaviorism, but machine learning is not behaviorism (Domingo, 36). Today, learners are being used by people from all walks of life and as our technology grows, so will our dependency on automation. It's remarkable to see just how much data science, machine learning, and artificial intelligence influence the world around us. Even though bridging the gap between far-fetched ambitions and a world virtually run by automation remains vast, it will undoubtedly be interesting to see how much that gap closes in the future.

Works Cited

Domingos, Pedro. (2015). *The Master Algorithm*. Philadelphia, PA: Basic Books.

Stephens-Davidowtiz, Seth. (2017). *Every Body Lies*. New York, NY: HarperCollins Publishers Inc.

Tegmark, Max. (2017). *Life 3.0*. New York, NY: Penguin Random House LLC.

Wikipedia contributors. "Probably approximately correct learning." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 22 Jan. 2019. Web. 7 Mar. 2019.