

SUMMARY

THE MASTER ALGORITHM

PEDRO DOMINGOS



 **QuickRead**
FREE BOOK SUMMARIES

Summary of “The Master Algorithm” by Pedro Domingos

Written by Lea Schullery

How the Quest For the Ultimate Learning Machine
Will Remake Our World.

Introduction	6
Machine Learning Could Answer All of Our Future Problems	8
The Power of Algorithms and How to Prevent Them From Finding Too Many Patterns	10
Decision Trees Can Be Used to Prevent Overfitting and Are Incredibly Accurate	12
Simplifying Assumptions Prevent Accurate Algorithms From Overfitting	14
Clustering Algorithms and Neural Networks are Effective Ways to Sift Through Data	16
The Business of Buying Data	18
The Master Algorithm and Your Digital Mirror	20
Final Summary	22

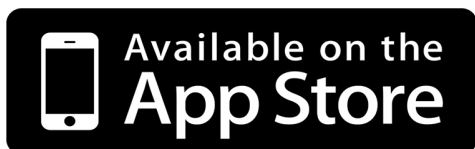


QuickRead

F R E E B O O K S U M M A R I E S

Go to QuickRead.com/App now to download our app and get access to thousands of free book summaries as both text and audiobooks.

Get the key insights of non-fiction books in minutes instead of hours. Listen to our free audiobooks while you workout or on your commute to work.



Introduction

Machine learning is everywhere nowadays, even in places you don't recognize! For instance, when you type a query into a search engine, machine learning is how the engine figures out which results to show you. When you open up your email, you don't see the spam because machine learning has filtered it out. When you go to Amazon.com or Netflix, a machine-learning system is recommending products or movies you might like. Even social media sites like Facebook and Instagram use machine learning to decide whose profiles you'll see and which updates you'll read. Today, machine-learning algorithms, also known as *learners*, are able to make inferences about what you want to see and products you want to buy by making inferences from data. The more data they have, the better they get! Machine learning is found in more than just the computer; in fact, your entire day is likely saturated with it. Think about it. Perhaps your clock radio goes off at 7:00 a.m. and is playing a playlist from Pandora. As you listen to music through the app, Pandora continues to learn your tastes in music thanks to machine learning. Then, as you drive to work, your car continues to adjust fuel injection and exhaust recirculation to get the best gas mileage. You can even use an app like Google Maps to predict traffic on your commute.

Your cell phone itself is full of learning algorithms. It corrects your typos, understands your spoken commands, recognizes bar codes, and so much more. At work, your email sorts itself into folders, leaving only the most important messages in your inbox. Your word processor checks your grammar and spelling. When you head to the supermarket after work, the aisles you walk down are laid out in a specific way with the help of learning algorithms. It determines which goods to stock and which end-of-aisle displays to set up. Paying with a credit requires a learning algorithm to determine whether or not you can be approved. When you get home, the letters in your mailbox are routed to you by a learning algorithm that can read handwritten addresses. Your junk mail is selected for you by other learning algorithms. These examples are only the beginning. Technology is

continually changing, making machine learning more sophisticated than ever, perhaps even becoming as smart as the human brain. So if you're ready to explore the captivating world of algorithms and machine learning, then let's begin.

Machine Learning Could Answer All of Our Future Problems

Today, algorithms have been woven into the fabric of everyday life and they aren't just used in your cell phone and laptop anymore, they're used in your car, your house, your appliances, and more. They schedule flights and then fly the airplanes. They run factories, trade and route goods, cash the proceeds, and keep records. If every algorithm suddenly stopped working, our world would follow and stop working too. But what is an algorithm anyway? It's a sequence of instructions telling a computer what to do.

The computer is made up of billions of switches called transistors, algorithms then turn those switches on and off billions of times per second. By flipping a switch, transistors communicate with one another using logical reasoning. For example, if transistor A only turns on when transistors B and C are both on, then it's doing a tiny piece of logical reasoning. If A turns on when either B or C is on, then that's another tiny logical operation. And if A turns on when B is off, and vice versa, that's the third operation. In the end, every algorithm, even the most complex ones, can be reduced to three operations: and, or, and not. When we combine these operations, we can begin to carry out elaborate chains of logical reasoning.

An algorithm, however, isn't just a simple set of instructions. They have to be both precise and unambiguous. For example, when you follow a recipe to bake cookies, you aren't using an algorithm. Anyone who has followed a cookie recipe knows that the result might be delicious or a complete disaster. On the contrary, an algorithm always produces the same result. Additionally, every algorithm has an input and output; that is, they are designed to produce a result based on the information it's given. But machine learning does more. They are given information as input and produce another algorithm as the output!

In other words, computers can write their own programs. In fact, computers can learn programs that people can't even write. For instance, when we decipher handwriting, we do so unconsciously, you can't quite put into words exactly the process you go through. Luckily, with machine learning, you don't have to. You can simply give a machine learning algorithm examples of handwritten text as input, and the meaning of the text as the desired output. The algorithm will then transform one into the other! Once learned, the algorithm can be used whenever you want to decipher handwriting automatically. This is exactly how the post office reads zip codes and why self-driving cars are on the way.

Machine learning requires just a few things, and one of the most important is data. The more data, the more it can learn. Given enough data, a learning program can solve almost any problem. Even more, machine learning can use the same algorithms to solve various non-related problems. Typically, if you have two different problems to solve, you need to write two different programs. Machine learning is different. But can it be possible for one learner to do everything? Could a single algorithm learn all that can be learned from data? Is a Master Algorithm the answer to all our problems?

The Power of Algorithms and How to Prevent Them From Finding Too Many Patterns

With all the research we have on machine learning, scientists have identified major “tribes,” where each one approaches problems differently and is defined by a set of core beliefs. The first is the symbolists, who believe intelligence can be reduced to manipulating symbols, much in the same way mathematicians solve equations by replacing expressions with other expressions. They also understand that you can’t learn from scratch; instead, you need some initial knowledge to go with the data. Therefore, symbolists have figured out how to incorporate preexisting knowledge into learning and how to combine different pieces of knowledge to solve new problems.

The key influencer of the symbolist tribe is David Hume, one of the greatest empiricists and English-speaking philosophers of all time - “the patron saint of the symbolists.” Born in Scotland in 1711, Hume spent much of his time debunking the myths of his age. He asked the profound question, “How can we ever be justified in generalizing from what we’ve seen to what we haven’t?” In a sense, every learning algorithm is an attempt to answer this question. 250 years after Hume asked this question, physicist David Wolpert created the “no free lunch theorem,” which essentially sets a limit on how good a learner can be.

In machine learning, you need both positive examples and negative examples. For instance, if you’re trying to learn to recognize cats in images, images of cats are positive examples, and images of dogs are negative ones. These learning algorithms, however, are prone to overfitting because they have an almost unlimited capacity to find patterns in data. And if you search enough, you can find anything which is both its strength and its weakness. For example, In 1998, *The Bible Code* claimed that the Bible contained predictions of future events that you could find by skipping letters at regular intervals and assembling words from those letters. Critics demonstrated that these predictions can be found in almost any long text

and proved their argument by finding patterns in *Moby Dick* and Supreme Court rulings.

This is an example of overfitting, or hallucinating patterns, which takes place when an algorithm is so powerful that it can learn anything. Put more simply, a data set that is as large as the Bible will almost always produce patterns. The key is to get your algorithms under control by limiting their complexity and placing restrictions. With the right restriction, you can ensure the range of your algorithm isn't too wide, leading to consistent results. On the other hand, if it's too flexible, you might end up with something like *The Bible Code* that finds patterns in any given set of data. But how do you decide whether to believe what a learner tells you?

The key is that you don't believe anything until you've verified the results. To do this, "you take the data you have and randomly divide it into a training set, which you give to the learner, and a test set, which you hide from it and use to verify its accuracy. Accuracy on held-out data is the gold standard in machine learning."

Decision Trees Can Be Used to Prevent Overfitting and Are Incredibly Accurate

Symbolists are the oldest branch of the AI community. They are rationalists and therefore rely on logical methods for intelligence. For this reason, symbolists prefer inverse deduction algorithms. Inverse deduction creates rules by linking separate statements. For example, if you have the statement, “Socrates is human” and “Therefore Socrates is mortal,” the algorithm can arrive at broader statements, like “All humans are mortal.” On the other hand, we don’t induce that all mortals are human because there are many other mortal creatures, like cats and dogs.

Today, inverse deduction plays an important role in predicting whether new drugs will have harmful side effects. When you can generalize from known toxic molecular structures, you can form rules that quickly weed out many apparently promising compounds, which greatly increases the chances of successful trials on the remaining ones. Furthermore, with this knowledge, we can begin to predict which drugs will work against cancer genes. While this sounds promising, there are still many limitations when it comes to inverse deductions; one of those limitations being that it is costly and inefficient to work with massive data sets.

For these, the symbolist algorithm uses a decision tree induction. Decision trees use a “divide and conquer” algorithm to branch off the data into smaller sets. Essentially, they play a game of 20 questions to further narrow the options and possibilities. The first step is to pick an attribute to test at the root, like testing to determine if you are a Republican, Democrat, or Independent. From there, you will focus on examples that go down each branch and then pick the next test. One example would be checking whether tax-cutters are pro-life or pro-choice. You then repeat this for each new branch until all the examples have the same class, at that point, you simply label the branch with the class you are testing: Democrat, Republican, or Independent.

Decision trees are a great method to use to prevent overfitting. By restricting the number of questions the decision tree asks, only the most widely applicable and general rules can be applied. We see decision trees often in software that makes medical diagnoses by narrowing down a patient's symptoms. Additionally, decision trees can be far more accurate than humans. For example, in 2002, a head-to-head competition revealed that decision trees correctly predicted 75% Supreme Court rules, while a panel of experts got less than 60% correct.

Simplifying Assumptions Prevent Accurate Algorithms From Overfitting

Another popular tribe and branch of machine learning is Bayesianism. Bayesians are empiricists, which means they believe all reasoning is fallible and that knowledge must come from observation and experimentation. They believe all learned knowledge is uncertain, and learning itself is a form of uncertain inference. The problem then becomes how to deal with noisy, incomplete, and contradictory information. The answer for Bayesians is Bayes' theorem, which tells us how to incorporate new evidence into our beliefs.

This approach works particularly well for medical diagnoses. For instance, if you test positive for AIDS, then your probability of having it goes up. When you add the results of multiple tests, combining them would result in a combinatorial explosion; therefore, you'll need to make simplifying assumptions. Furthermore, you must consider many hypotheses at once, such as all the different possible diagnoses for a patient. Computing the probability of each disease from the patient's symptoms can take a lot of time, so here is where learning the Bayesian way comes in handy. By using a simple cause and effect formula, Bayesian inference becomes a powerful algorithm.

For example, when making a flu diagnosis, a doctor doesn't do so based on a single symptom, like having a fever. Instead, she takes many symptoms into account, including whether you have a cough, sore throat, runny nose, headache, chills, and more. Therefore, the doctor makes simplifying assumptions that whittle the number of probabilities down to something that is much more manageable. That restricting assumption is to assume that two symptoms do not influence one another, meaning a cough doesn't affect your chances of also getting a fever. In this way, Bayesian inference avoids overfitting by strictly focusing on the connection between cause and effect.

Assumptions like this are also used by voice-recognition software like Siri. Imagine you just said “Call the police,” Siri then considers the probability of you saying “tell” instead of “call” or “please” instead of “police.”

Individually, the most likely words are likely *call*, *the*, and *please*. But the sentence “call the please” forms a nonsensical sentence, so taking the other words into account, Siri concludes that the sentence is really “Call the police” and makes the call.

Clustering Algorithms and Neural Networks are Effective Ways to Sift Through Data

When you become a parent, you suddenly see the mystery of learning unfold before your eyes in the first three years of your child's life. A newborn baby can't talk, walk, or recognize objects. But month after month, the child continues to take small and large steps towards figuring out how the world works. And by the child's third birthday, all this learning has created a stable self in which the mind has grown into a stream of consciousness that will continue throughout life. It is this phenomenon that brings us one step closer to the Master Algorithm.

As we grow up, we organize the world into objects and categories. What we need is an algorithm that will spontaneously group together similar objects, or different images of the same object. These types of algorithms can be used in image recognition or voice isolation software, which identifies a face or object among millions of pixels. Likewise, each thing you click on Amazon provides a sliver of information about you. Little by little, all those clicks add up to form a picture of your taste, much in the same way all those pixels add up to a picture of your face.

You see, a face has only about fifty muscles. The shape of the eyes, nose, mouth, and so on only have about 10 different variations for each feature, which is how a police sketch artist can put together a sketch of a suspect that's good enough to recognize him. By memorizing those ten different variations, they narrow their options down and make it possible to produce a drawing based on a single description. Similarly, facial recognition algorithms only need to compare a few hundred variables versus a million pixels.

Lastly, another effective tool for sifting through large amounts of raw data is neural networks. Through the use of stacked autoencoders, learners can work as a brain and process multiple inputs at the same time. Through nonlinear neurons, each hidden layer learns a more sophisticated

representation of the input, building upon the previous one. For instance, “when given a large set of face images, the first layer learns to encode local features like corners and spots, the second uses those to encode facial features like the tip of the nose or the iris of an eye, the third one learns whole noses and eyes, and so on. Finally, the top layer can be a conventional perception that learns to recognize someone, let’s say your grandma, from the high-level features provided by the layer below it.”

One of the biggest neural networks created was The Google Brain network, which consisted of a nine-layer sandwich of autoencoders that learned to recognize cats from YouTube videos. With one billion connections, it was the largest network ever learned at the time. One of the project’s principals, Andrew Ng, is one of the leading proponents of the idea that human intelligence can be boiled down to a single algorithm, and we just simply need to figure it out! These stacked autoencoders are just one step closer to solving AI.

The Business of Buying Data

Why is Google worth so much more than Yahoo? If you look at what they're doing, they both are essentially doing the same thing. They both show ads on the web, they're both top destinations, and they both use auctions to sell ads and machine learning to predict how likely a user is to click on an ad. But there is one difference. Google's learning algorithms are much better than Yahoo's. Today, the company with the best algorithms is the company that's going to be the most successful.

In every market, producers and consumers need to connect before a consumer is willing to make a transaction. In pre-Internet days, the obstacles that companies encountered were physical, like not having enough shelf space for new products. Today, the problem becomes the overwhelming number of choices. How do you browse the shelves of a bookstore that has millions of titles for sale? We can apply this to all other goods as well, like shoes, hotel rooms, investments, even music, news, and videos. It even applies to those looking for a job or a date. How do you find each other?

Machine learning helps narrow our choices down. Amazon, for example, offers suggestions on what products customers might like based on clicks. Even better, their service covers just about every market you can think of. While Amazon leads the race in learning algorithms, there's a mad race to gather data about you. Everybody loves your data, it's the gateway to your world, your money, your vote, and even your heart. But everyone only has a small portion of it. Google sees your searches, Amazon sees your purchases, AT&T your phone calls, Capital One your credit-card transactions, and more. No company has a complete picture of you, which is both good and bad. Clearly, it's good because if someone knows *everything* about you then they would have far too much power. It's bad, however, because there will never be a 360-degree model of you.

Your data can be a tremendous asset to a company, which is why there is value in your data. Today, the value of a user to the Internet advertising industry is around \$1,200 a year. Google's sliver of your data is worth about \$20 while Facebook's is worth \$5. As data increasingly becomes the new oil, some existing companies would love to host the digital you. However, companies like Google and Facebook are not the best companies to house your digital self because of their conflict of interest: targeted ads. Both companies use targeted ads and so would have to balance your interest and the advertisers'.

The solution to this problem is databanks who keep your information secure and allow you to determine when and how it is accessed. Think about labor unions at the start of the twentieth century which were started to balance the power of workers and bosses. The 21st century needs data unions for a similar reason. A data union would allow its members to bargain on equal terms with companies about the use of their data. Today, most people are unaware of how much data there is about them and how it's being used. Even more, they don't know what the potential costs and benefits are and companies are simply flying under the radar, using your data how they want, terrified of a "blowup." A blowup will happen, so it's best to raise awareness now and let everyone make their own decisions about what they wish to share, and how and where it's used.

The Master Algorithm and Your Digital Mirror

If you're thinking there can't be that many companies with data on you, then let's take a look at all of the places where your data is recorded: your emails, office documents, texts, tweets, Facebook and LinkedIn accounts, web searches, clicks, downloads, and online purchases; your credit, tax, phone, and health records; your Fitbit statistics, your driving that is recorded by your car's microprocessors; your wanderings recorded by your cell phone; all the pictures of you ever taken, including brief cameos on security cameras, and so on. Your data is everywhere, and when combined, this data could form a fairly accurate and detailed picture of you. But what could this mean for your future?

No learner today has access to all this data, and even if it did, we still can't be sure how accurate it would be. But suppose you took all that data and gave it to a very real, future Master Algorithm. In this case, the Master Algorithm could learn a model of you, and you could essentially carry that model in a thumb drive in your pocket to be used as you wish. This digital mirror could show you things about yourself that you don't yet know. It would have the power to suggest just a dozen books from the world's marketplace. It could do the same for movies, music, games, clothes, electronics, and more. No longer would you have to search endlessly for items you wish to purchase, this digital mirror could suggest items based on your activity.

Additionally, it could keep your refrigerator stocked at all times, filter your email, voice mail, Facebook posts, and Twitter feed, and even appropriately reply on your behalf. It would even remove little annoyances in your life, like checking your credit card statements, disputing charges, renewing subscriptions, and filling out tax returns. It would find a remedy for your ailment, run it by your doctor, and order what you need from the local pharmacy. It would even suggest interesting job opportunities and vacation spots, and suggest potential candidates to vote for on the ballot. And if

you're single, it could find you a potential date. After selecting a match, it would communicate with your date's digital model and pick restaurants that you both might like.

In the future, everyone could have another "digital half" communicating with the digital selves of others. For instance, if you're looking for a job and a particular company is interested in you, the company's model will interview your model, similar to a face-to-face interview - but it will only take a fraction of a second. You'll be able to click on "Find Job" in your future LinkedIn account, and you'll immediately interview for every job in the universe that fits your parameters, such as profession, location, and pay. In the world of the Master Algorithm, "Have your people call my people" becomes "Have your program call my program." In the future, your digital self will be able to do everything you don't have to, making your life easier and more efficient.

Final Summary

Now that you know the secrets of machine learning, you now understand how data is turned into knowledge. You understand what it can and cannot do and the underlying complexities it brings. You know what companies like Google, Facebook, Amazon, and the rest do with the data you generously give them every day and why they can easily find stuff for you, filter out spam, and keep improving what they offer you. But we still don't have everything figured out, that's because a Master Algorithm doesn't yet exist. We only know what it could look like. But the first step is knowing the possibilities. Perhaps something is missing, something that hasn't been discovered yet. This is where you come in. It's up to us to ensure the Master Algorithm falls into the right hands, which is why it should be open-sourced and shared with everyone on Earth. Lastly, "Newton once said that he felt like a boy playing on the seashore, picking up a pebble here and a shell there while the great ocean of truth lay undiscovered before him. Three hundred years later, we've gathered an amazing collection of pebbles and shells, but the great undiscovered ocean still stretches into the distance, sparkling with promise. The gift is a boat - machine learning - and it's time to set sail."



QuickRead

F R E E B O O K S U M M A R I E S

Go to QuickRead.com/App now to download our app and get access to thousands of free book summaries as both text and audiobooks.

Get the key insights of non-fiction books in minutes instead of hours. Listen to our free audiobooks while you workout or on your commute to work.

