The Master Algorithm - Pedro Domingos

Notes from "The Master Algorithm" - Pedro Domingos

(Ref: Pedro Domingos: "The Master Algorithm" | Talks at Google)

The book is ...

- All about"Machine Learning"
- Today there are five tribes (as categorized by Domingos), each with their own master algorithm.
- Best for itself but not for others.
- The ultimate master algorithm combines them into a general purpose learning machine.

The Origins of Human Knowledge

- Evolution: Knowledge encoded in our DNA that makes us what we are. Built through Survival of Fittest.
- Experience: Acquired by living in the world.
- Culture: from culture, from talking with other people, from reading books and so on.

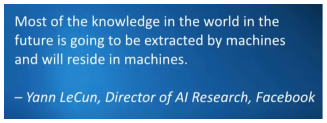
These are the sources of knowledge in natural intelligence. The next one far better than the previous one.
There is a new source of knowledge on the planet and that's ...

(Ref: Shane Parrish Blog)

The Origins of Human Knowledge

Computers ...

- Discovering knowledge from data.
- for example, these days there are hedge funds that are completely run by machine learning algorithms.
- They look at the data, they make predictions and they make buy and sell decisions based on those predictions.



5 main paradigms in Machine Learning

COmputers discover knowledge by mimicking following processes:

- Filling Gaps: Scientists put forth hypothesis/theories, make observations, adapt the theory to fill the gap.
- Reverse Engineering: Emulate the brain. The greatest learning machine on earth is under your skull.
- Simulate Evolution: More superior than brain, as it creates brain, body, other life forms, etc.
- Quantify uncertainty/probability: As more evidence comes, adjust it, using Bayes' theorem.
- Reason by analogy: You try to find matching similar situation from past, apply its solution.

For each of these, there is a school of thought ...

5 Tribes

- Symbolists: Learning as the inverse of deduction
- Connectionists: Reverse engineer the brain, seen in back-propagation
- Evolutionaries: Simulate evolution on the computer, seen in genetic programming
- Bayesians: A form of probabilistic inference
- Analogizers: Extrapolating from similarity judgments, seen in Support vector machines.

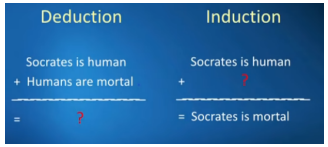
Each of these have their own master algorithm, to learn ANLYTHING (given enough data, proof available).

Symbolists

- Filling in the gaps of the knowledge we have
- Believe in the power of logic
- Master Algorithm: Inverse deduction
- Famous folks:

Inverse Deduction

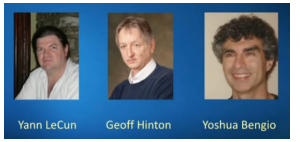
- Learning is induction of knowledge
- Deduction is "general rule to specific facts"
- Induction is "specific facts to general rule"
- Example: Addition: If I add 2 and 2, what do I get? : 4
- Subtraction gives answer to the inverse question: What do I add to 2 to make it 4? : 2
- Similarly, inverse deduction gives, what else do you need to know, to come to the answer
- A general rule : "Humans are mortal"



Rules can be composed into much more complex system to match multitudes of answers.
Thats the Machine Learning!! finding the formula/function/rule given inputs and output.

Connectionists

- Take their inspiration from the workings of the brain. Neural Networks.
- Master Algorithm: Backpropagation.
- Famous folks:



Working of Neurons

- A neuron gets various inputs. If the weighted total crosses the threshold, it fires, and sends electrochemical signal downstream to the connected neurons.
- The connections are called, synapse. Stronger they are more weights, threshold crossings are more, more transmissions happen, more learning.
- What we thought is not the actual, we correct our understanding, thats how we learn.
- Same thing happens as Back-propagation, which adjusts the weights. So that only correct neurons (branches) fire.

Thats Deep learning!!

Evolutionaries

- Learn the structure itself. Not need to feed it like Neural networks do.
- Test learning algorithms by their "fitness", a scoring function as to how well the algorithm meets its purpose.
- The successful algorithms are then mutated and recombined (sex) to produce new algorithms that can continue the competition for survival.
- Eventually an algorithm will find a fitness peak where further mutations or recombination do not increase the algorithm's success.
- Master Algorithm: Genetic Programming
- Famous folks:

Thats Genetic Programming!!

Connectionists vs Evolutionaries

- Neural networks start with a predetermined structure. Genetic algorithms can learn their structure (although a general form would be specified).
- Backpropogation, the staple process by which neural networks are trained, starts from an initial random point for a single hypothesis but then proceeds deterministically in steps to the solution.
- A genetic algorithm has a sea of hypotheses competing at any one moment, with the randomness of mutation and sex potentially producing big jumps at any point, but also generating many useless algorithm children.

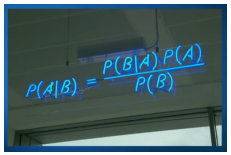
(Ref: Jason Collins Blog)

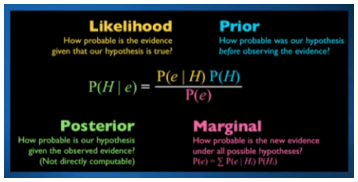
Bayesians

- Start with a set of hypotheses that could be used to explain the data, each of which has a probability of being true (their 'priors').
- Those hypotheses are then tested against the data, with those hypotheses that better explain the data increasing in their probability of being true, and those that can't decreasing in their probability.
- This updating of the probability is done through Bayes' Rule.
- Famous folks:

What is Bayesian Learning?

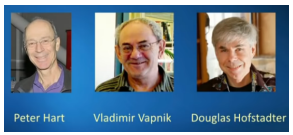
- Everything that you learn is uncertainty/probability. Compute the probability of your hypothesis and update it as the new evidence comes in using Bayes theorem.
- Master Algorithm: Bayes Theorem
- Likelihood: If your hypothesis is such that it makes observations likely, then conversely, what you are observing is what makes your hypothesis likely.



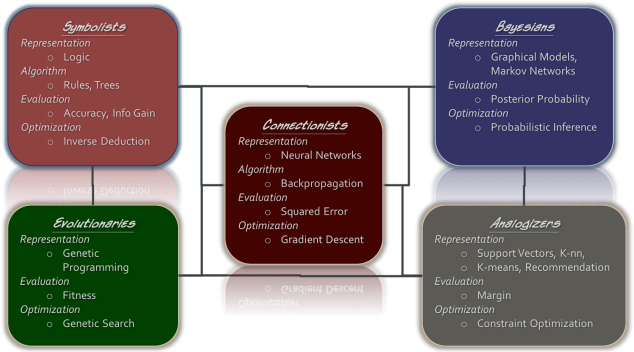


Analogisers

- Reason by analogy:apply solution of the problem in past, that matches with the current ones.
- Master Algorithm: KNN, Support Vector Machines.
- KNN can form a complicated partition. One interesting fact that, points, away from partitions are irrelevant. Only near the boundary control it. These are same as Support vectors. Same as SVM with Kernel trick.
- Famous folks:



Summary of Tribes



Summary of Tribes

Tribe[5]	Representation[6]	Evaluation[7]	Optimization[8]
Symbolists	Logic	Accuracy	Inverse Deduction
Connectionists	Neural Networks	Squared Error	Gradient Descent
Evolutionaries	Genetic Programs	Fitness	Genetic Search
Bayesians	Graphical Models	Posterior Probability	Probabilistic Inference
Analogizers	Support Vectors	Margin	Constrained Optimization

translated from figure on p. 240.

But we need a single algorithm that solves all of these …
Thats the Master Algorithm, a Grand Unified Theory of Machine Learning.

What is The Master Algorithm?

"a general-purpose learner" (p. xxi) …an algorithm that, "if it exists, [it] can derive all knowledge in the world—past, present, and future—from data" (p. xviii) …an algorithm that can "learn to simulate any other algorithm by reading examples of its input-output behavior" (p. 34) …"the unifier of machine learning: it lets any application use any learner, by abstracting the learner into a common form that is all the applications need to know" (p. 237).
What a promise!!!

How can we combine all these algorithms?

They all have same three parts:

- Unifying Representation: Markov Logic networks, rules with probability/weights.
- Evaluation: A scoring function that declares how good the model is. How does it fit data? How well it answers accurately? Given by the consumer, we need to optimize it.
- Optimization: How to optimize the model/function. Use Genetic process or crossing, come up with different combination and see which one works.