One learning algorithm

Abhi Agarwal (abhia@nyu.edu)

The one learning algorithm is a concept in machine learning and neuroscience that has been studied extensively by Jeff Hawkins. In his book, On Intelligence, he laid out a memory-prediction framework depicting what he thinks the one learning algorithm would consist of. His framework is based upon the work on columnar organization of the cerebral cortex done by Vernon Mountcase who shows that “all regions of the cortex [perform] the same operation” (Hawkins, 51), and that “there is a common function, a common algorithm, that is performed by all the cortical regions” (Hawkins, 51). In essence, the one learning algorithm accurately predicts outcomes of future input sequences based on patterns that this algorithm learns from regardless of what type of patterns they are (patterns can be images, videos, smell, etc.).

Hawkins is an electrical engineer, and hasn’t had any professional experience in neuroscience. His framework approaches the problem from an engineer’s perspective as well as his personal study of the research done on the cerebral cortex to formulate his framework. He treats the network of neurons in our brain as an encoding problem, and he quotes “All your brain knows is patterns” (Hawkins, 56), which is truly how a mathematician would see it as patterns imply that some formulation could be found to connect the pattern together. He explains that our understanding of the world is based upon patterns and “[correct] predictions result in understanding. Incorrect predictions result in confusion and prompt you to pay attention’ (Hawkins, 89).

Hawkins key aim is to create an algorithm that mirrors how we learn, and form what he calls 'true intelligence' rather than forming 'artificial intelligence'. He defines true intelligence as a "[measure of the] capacity to remember and predict patterns in the world, including language, mathematics, physical properties of objects, and social situations" (Hawkins, 97). Hawkins' view of how our brain recognizes patterns and learns is that our "brain doesn't "compute" the answers to problems; it retrieves the answers from memory ... entire cortex is a memory system. It isn't a computer at all" (Hawkins, 68). Thus, Hawkins sees the brain as being able to predict the answer or the best outcome of a certain task by having a very efficient prediction system that utilizes the connection of neurons in our brain to find the best prediction. This is in contrast to how most individuals think about learning and how the neurons in the brain responses to actions. Most individuals believe that the reactions to actions are, in a sense, computed at that particular moment and our brain is built on logic, assumptions, and prior knowledge. This contrast arises the question is learning something we can model as patterns? Can how we think be quantified and put into a series of steps or is it more complex? Can the one learning algorithm be used to represent 'true intelligence', and the process of learning in a similar way that our brains do?

Approaching this algorithm from an mathematical perspective, a mathematical model of learning would ignore a lot of things that we believe contribute to learning, and our definition of learning. Lets take an objective definition of learning to be the "process of acquiring modifications in existing knowledge, skills, habits, or tendencies through experience, practice, or exercise" (Merriam-Webster, "Learning"). Knowledge to us is an understanding of pieces of information through experience or learning, and we usually recall these pieces of information using languages that we know, or visual descriptions that we collect. However, to a computer knowledge is represented as a series of numerical values and things a computer learns are converted, through some measure, into numerical values to be processed through the algorithm and then stored. Without representing them as numerical values there is no way a computer can understand their meanings, as these algorithms require some medium to compare and relate two pieces of information and to classify this new piece of information. Once we input a phrase into this particular learning algorithm it becomes arbitrary. It gets converted into a series of values that the computer performs actions upon to match the particular pattern, and to try to classify it. In contrast, our brains understand this information as it has complex biological traits that allow it to interpret and break down these words, and processes them through its existing knowledge. The question arises that can a learning algorithm react to situations using true intelligence?

If we define true intelligence to be "the ability to learn or understand things or to deal with new or difficult situations" (Merriam-Webster, "Intelligence") then we can begin to compare it to Hawkins definition of true intelligence, and evaluate if his definition can be used to represent true intelligence. Just concentrating on Hawkins' definition of intelligence we can observe that it is a more mathematical approach to think about intelligence. Hawkins concentrates more on things that can be quantified such as capacity, and prediction of patterns whereas the other definition concentrates on the ability and the understanding. A learning algorithm doesn't have the ability to understand the information that's given, but only the ability to use it to predict using existing pieces of information that is most similar to it by some index.

Human traits such as instincts, courage, self-awareness, reasoning, planning, etc. would also be treated as patterns that can be detected by the algorithm, and then classified according to their actions if we adhered to Hawkins definition of intelligence. Reasoning is not a trait that a computer can begin to make use of by looking at patterns, and Hawkins argument was that truly intelligent machines would not have to have anything explicitly programmed within them (apart from the learning algorithm). It requires a sense of logic, and an understanding of how information relates to each other. Moreover, in order to train the algorithm to use these traits we would need to quantify them, and to break them down into formulations a computer can understand. The designer would have to lay out explicit steps or a clear boundary of when the algorithm would decide to either be courageous or to be not courageous, and explicit steps on using the information it has to create reasoned arguments. This would reduce our traits to being binary, either we are being courageous at a particular moment or we are not and the separation between those would be where a particular numerical value lies on a plane or on a scale. A particular being would have true intelligence if the intelligent being would be able to explain and understand the decision it is making, and for the intelligent being to show the logic that led him to that particular decision.

In addition, in the definition of learning there was an aspect of experience. What we perceive as experience and what the one learning algorithm would define to be experience would be very different. Experience is central to learning, and in regards to how we perceive learning, experience is knowing the prerequisites to the topic at hand or having had experience and some existing knowledge on the topic. To Hawkins' one learning algorithm experience would be defined as a numerical value such as time, or number of connections to other existing knowledge, or time spent to add and update a particular piece of information. His representation of intelligence is having a large capacity to store these pieces of information, and then having algorithms to predict other pieces of information given these. In this way there is no other way to represent experience as it's hard to understand where it belongs in your knowledge.

Moreover, it's quite simple to think about creating learning algorithms to target specific traits of human intelligence and imagining them being able to closely mirror true intelligence, but Hawkins defines a one learning algorithm. Creating specific learning algorithms would imply that we would be creating algorithms that would be targeted to learning these traits. We would be able to take certain patterns and predict if we need to be courageous or not, and therefore we would be able to define certain aspects that correspond to them. This wouldn't be a learning algorithm that would mirror true intelligence, but it would be an algorithm that would be the closest to mirror true intelligence. However, the one learning algorithm would take any set of patterns and make a prediction based on the existing information it has regardless of if it is a trait or a piece of information. This brings in issues as each particular prediction is based on quantified attributes that you define, and you have to define these particular attributes to fit everything you're trying to represent. This is in contrast to the specific learning algorithms where you're able to define attributes that directly correlate to the data.

In conclusion, given Hawkins definition of intelligence the learning capabilities of a computer could never will truly intelligent, but be artificially intelligent. We can create separate learning algorithms that are able to mirror or get close to true intelligence by providing large training sets, and having fixed attributes. We would develop these attributes by quantifying learning and intelligence and using those arbitrary values to make predictions about the future.

References

[1] Hawkins, Jeff, and Sandra Blakeslee, *On Intelligence*. New York: Henry Holt, 2005. Print.

[2] Merriam-Webster, *Learning*. Merriam-Webster, n.d. Web. 05 Nov. 2014. Electronic.

[3] Merriam-Webster, *Intelligence*. Merriam-Webster, n.d. Web. 05 Nov. 2014. Electronic.