

In Our Image

“We are special. We are different. We are human.” To argue about the innate conceit that has dominated humanity’s view of itself with relation to the universe would to this day be incendiary in many circles, and in the end, entirely subjective and fruitless. This mindset, while it has been jarred constantly by scientific discoveries, from Copernicus and the non-Geocentric universe to the recent mapping of the human genome, remains a strong influence on the way people view the world. While this view may or may not be reasonable and productive, there is one particular area in which it is present that would make for interesting study. A field whose very name could be read to suggest the view that nature, and indirectly humans, are perhaps not superior to, but certainly separable from, that which is fabricated: *artificial* intelligence.

When a computer science professor asked one day during lecture for students’ ideas of what artificial intelligence is, one student suggested that it is a machine in which every input and circumstance has been foreseen and an appropriate action, consistent with human behavior, has been chosen as a response. While traditional AI has tried this approach, and in fact algorithms like this do exist, such as Markov Decision Processes, as well as real life examples like the famous Deep Blue which was able to play and win chess games on a world championship level, this view of AI is a popular misconception which trivializes the field’s very existence. And some, such as Hubert Dreyfus in his book *What Computers Still Can’t Do*, indeed offer this very suggestion: that the field is inherently flawed. But critics and those who suggest tests, both in discourse and practice, for the efficacy of AI all have a tendency to use a single benchmark: the human mind. This is apparent in the formulation of the well-known Turing Test, proposed by Turing as a possible way to test AI by allowing a human to interact with a computer by typing. A computer would pass the test if the individual conversing with it could not distinguish between the computer and a real human being. But often within this sort of hypothetical, informal discourse about the outwardly apparent characteristics of artificial intelligence, the true meaning the pursuit holds for us is lost among the vague attempts at definitions and contexts for intelligence based on our own human experience. As a more practical response to this human-centered mindset, AI has developed a branch which deals with neural networks made in our brain’s image and attempts to replicate our brain’s functionality by emulating at a very low level. While this attempt to model our behavior has had moderate success (by contemporary standards) as an approach to artificial

intelligence, its shortfalls have at the same time allowed us to re-evaluate our conceptions and our benchmarks for intelligence. This reassessment can take different forms, however, from Dreyfus's slightly outdated and but still grave indictment of traditional AI approaches, to the less philosophical and more deterministic comparisons that can result from these more discrete, information-based approaches to AI. By looking at neural networks (and a few other pertinent examples) with Dreyfus's criticisms in mind, we will find that this computational attempt to solve the AI problem allows us to look at intelligence in a way that may not completely satisfy critics like Dreyfus, but is useful nonetheless.

As a field of study, AI focuses primarily on learning, something viewed by the field as an exercise in computation, and while critics like Dreyfus may disagree with this view (Zaft para 4), it is, in a sense, built into the field due to the fact that attempts to create AI all utilize a computing platform, or more specifically, the computer. This driving goal, learning, seen as a problem of pattern recognition, is the very function performed by neural networks. A newer set of data structures, neural networks were developed informally in the 1950s by Minsky but formally introduced in the 1980s and based on the structure of the biological network of neurons present in our brain (Pfeffer "Perceptrons" 2). They represent information in as indiscrete a fashion as real numbers will allow (thresholds and weights are of varying continuous values, as opposed to discrete classifications). Neural networks allow for reinforcement and supervised (and sometimes even unsupervised) learning, the recognition of higher-level features of any relative scale in even the most minutely detailed data, and a method of recognizing these details which is free from any constraint if need be. As such, it is a method of storing observations and data that are not discrete, and in fact is very flexible.

Neural networks share three characteristics with our brains that make them valuable as learning algorithms. They have the potential to process information in parallel, given the multiple processor architecture to do so. They exhibit graceful degradation, meaning a neural network's performance will not dramatically decrease in any area if pieces of it are cut off. And finally, they have the ability to adapt based on experience (Pfeffer "Perceptrons" 4). Neural networks have been successfully applied in a variety of ways. In 1987, NetTalk was able to learn how to phonetically pronounce English words at a 95% accuracy rate on its training data, and 78% accuracy on its test data (Pfeffer "Neural Network" 6). Neural networks have successfully navigated vehicles across long stretches of highway, such as ALVINN, which was able to travel at 70 mph for 90 miles after being trained on images of a first person perspective of the road (Pfeffer "Neural Network" 7). A neural network was created in 1989 that could recognize handwritten zip code digits at a 98.9% accuracy rate (Le Cun, 7). Furthermore, neural networks are robust to noisy data and can deal with inputs for which they were not trained (and perform reasonably well on them). They resemble the human tendency to be versatile and yet slightly imperfect. The object of the neural net is in fact to be slightly

imperfect: the error function governing the back propagation algorithm is never meant to reach 0%, only its relative minimum. Most successes have been conspicuous, however, and architectures heavily constrained are the only ones that have been successful. Today, “researchers are unclear as to how well the method can be expected to perform in general...” (Pfeffer “Neural Network” 7).

One may become skeptical of neural networks’ abilities to model human-like behavior when this fact is introduced, but only if one ignores the fact that the human neural network is very likely fine-tuned and constrained itself. After all, humans have evolved over a very long period of time where the aspects necessary for our biological purposes were tuned to the actual surroundings and social structures, and very likely our brains are well-suited for their tasks, from abstract thought to spatial navigation to social communication. This long period of evolution is often ignored, and somehow AI researchers are expected to be able to fine-tune their algorithms over the period of a few months.

A more specific critique offered by Dreyfus is the idea of situational context which human beings have, a basis of understanding that allows them to learn and innovate in an informed and reasonable manner, and that implementations of existing AI algorithms cannot replicate this (Zaft para 6). But what is this vague notion of prior understanding? We can now try to define it, though our definition isn’t likely to be a satisfactory one. The first question to ask would be: where is it from? If we continue using a computational and somewhat deterministic paradigm, there are only two possibilities: it is either developed evolutionarily, or learned through observation and possibly reinforcement. This notion still seems almost inexorably vague, and more terms need to be defined. Specifically, we must try to define learning and innovation, and determine upon what, if anything, these rely.

There is an interesting set of theorems that most individuals, and even many mathematicians, will never learn about which are pertinent to this question; they are commonly known as the “No Free Lunch” theorems. One of them illustrates that in fact learning, in itself, is logically and mathematically impossible (Pfeffer “Decision Trees” 4). How is this so? It is easiest to explain by example: imagine there is a set of data available that says red light means off, and green light means on. You can observe this data, and even record it. But what happens next? If you subsequently see a red light, having learned what each means, can you say with any confidence that this time it in fact signifies off? Of course, you cannot. There is no reason to believe that the data is correct, and even if the data is based on prior experience, that it applies this time. This gets at one of the prime, if not the most important, aspects of learning: making assumptions. In fact, we have learned throughout our efforts to reproduce learning capacity that there are some very useful assumptions: an inductive bias allows us to apply past experience to future situations; a restriction bias allows us to find the simplest interpretation of observations. In fact, one need only look so far as the tried-and-true scientific method to find the value

of making assumptions to the ability to learn: every elementary school science class is taught that one makes a hypothesis, tests it, and either rejects or accepts the hypothesis based on observations. And would not this method work even if the assumptions you make are random? Certainly it would take time to learn, and yet it is conceivably possible to do so. It is in fact done so; in many learning algorithms, setting initial variables to random values results in better performance later on. Algorithms such as Markov Decision Processes and Q-Learning, which calculate the best action to take based on probabilistic analysis, reach higher levels of performance when seeded with sets of random numbers. In fact a whole computer science field has grown as a result, focused on producing random values of a better quality. So the point we should take away from all this is that to create anything new, anything beyond recorded observations, to learn or to innovate, one needs to add the novelty factor and make assumptions, random or "informed" (the effectiveness of reasoning also being a version of the inductive assumption).

But even if randomness is involved, humans tend to choose the correct span of random options or assumptions, they progress in their learning or innovation in a reasonable direction much of the time. The question then becomes, where does this propensity originate? If we assume that our brains are physical structures, there are, once again, only two possible sources: either the evolved characteristics innate in our neural structure, or the learning that took place within our lifetime. So likewise, in an AI algorithm, especially in neural nets, there are equivalents for both: constraints and assumptions built into the network, and training data which allows learning. We must remember, when talking about our own prior basis of understanding, that as infants and children we are also taught through reinforcement, and any learning we do on our own is initially thanks to "random" (perhaps evolutionarily hard-wired) exploration; one need only think of a baby, grabbing things, crawling all over the place, and babbling without any apparent (albeit possibly still present) purpose. Even though we begin with this random exploration, it still allows for reinforcement learning and later on what is learned from random exploration can inform and contribute to further learning and innovation. In fact these methods have been used successfully to create robots which explore and learn about a simple environment by spending a lot of time running into walls. So we see that neural nets actually counter Dreyfus's argument, if we take into account a more complete picture of human intelligence. But if neural networks are so promising and have existed for decades, why have they not been the silver bullet first envisioned by AI researchers?

The answer may lie in physical and temporal scales. Today, in 2004, a neural network composed of 246 neurons and 86,640 connections running on a fairly top of the line 3.2 GHz single processor (meaning each neural connection must be updated sequentially) will take about 40 seconds to update itself 10,000 times. An example of the type of function this kind of neural net may perform could be handwritten digit recognition. By comparison, the human brain is said to contain about 100 billion neurons

and about 1,000,000,000,000,000 connections, and updating is done in parallel at about 1 kHz (meaning any connection can be updated at the same time as another) over the period of a human lifetime (“Number of Neurons in a Human Brain”). Whether the brain has any innate specialized evolutionary characteristics which may improve its performance further, or whether its chemical composition influences the form in which information is represented is almost irrelevant when the gap in pure numbers is so great. Until processors can be built to match the capacity and parallel nature of the brain, questioning the usefulness of neural networks compared to humans is difficult, and any conclusions would be mostly conjectural. If a neural network contained as huge an amount of connections as a human brain, it would allow for the possibility of any set of sets of higher-level concepts which could all then be networked together to form even higher-level networks which react within the context of the outputs of all the lower level networks, from those that represent either memory, observations, or both in the form of neuron thresholds. Contextual recognition of relevant or irrelevant aspects of observations does not seem quite so impossible when one considers that “context” may merely be the summation of observations over a period of time. One need only calculate for themselves, using a concept called VC Dimension, that a single neural network composed of 300 (100 input, 100 interior, 100 output) neurons with at least a few interior layers can represent any continuous or discontinuous function, and can represent any hypothesis necessary to partition a set of items of size $(2 * 100 * 200 * \log_2(e * 200)) \sim 11656976$ (Pfeffer “Neural Network” 3). Imagine how many multidimensional relationships a human-sized neural network could represent. This shows once again that simply judging existing machine learning by human standards is impractical and unhelpful, given the reality of existing technology.

But even if the neural network were able to process information with a prior basis and context, would the information have *meaning* to the machine? This question, while surely worth asking, falls from the fallacious assumption that *our* form of “meaning” is most legitimate. It also ignores the fact that the observations which our senses provide us with are also meaningless until we adapt our neural network, as we grow, to the observations we perceive. This leads to one example, that of the human visual system. Blind individuals, born without sight, do not develop the same neural patterns in the visual-processing portion of the brain that individuals with sight do (Murray para 5). In fact, until our neural network adapts to the inputs offered by our eyes, nothing we see has any meaning, though it obviously has a purpose, being a source of input. In a more general approach, imagine an image of a cloud; it may have meaning to those of us who have seen them and were told what they are, but would it mean anything if we had never seen clouds before, or been taught to classify something like a cloud as a cloud? We may, perhaps, classify it based on its attributes; it is puffy, white, and soft-looking. The best we could do is to classify it based on its attributes and how its attributes compare to attributes of other things we see. In fact, is this not how we assign meaning to all our observations?

Learning algorithms such as Non-parametric Clustering do just this: they learn without supervision by simply clustering supplied data by common attributes (Pfeffer “Clustering Algorithms” 2). They do not need to be trained or instructed, information they receive does not need to be explained to them, and yet they assign meaning to it. Any subsequent data the algorithm is asked to classify it classifies using the clusters it found as it observed earlier data. It assigns its own “meaning” to data. The meaning behind information obtained through our senses, a slippery and vague sort of idea, can in the end only be judged in the context in which it can exist: our mind. It could simply be the “meaning” this information acquires through the adaptation of our neural network’s parameters to the information. This is a hard claim to make indeed, but one must admit that it is no more unreasonable than the claim that an artificial neural network’s adaptation of parameters (or that of any other algorithm with changing parameters, such as MDP, Q-Learning, Clustering etc.) to information is *less* “meaningful” within the context of *that* structure. We can only claim it is *different*.

So we see that the contrasts presented between humans and machines are either not quite so striking, or not quite that fair. Just because an algorithm cannot pass the Turing Test does not mean it is not as intelligent, or that its intelligence is somehow inferior, given its purpose and the context of its creation. To effectively judge existing networks and algorithms we need to judge them based on their contexts, which may be different from ours, or in the case of neural networks or clustering algorithms, let them make up their own. Helping them out by constraining their parameters is not necessarily cheating but quite probably very necessary, as well. Ultimately, AI theory is not so much at fault as impatience and a lack of sufficiently advanced physical engineering on our part.

As we look to ourselves to create effective AI methods, we find telling limitations which lead us to re-evaluate our perceptions of intelligence. This should show us that we should not necessarily try to make intelligence resemble us exactly. This is not the point. We should make intelligence for any number of purposes which we may find at a given time, and judge its merits as such, just like we judge the merits of our intelligence in the context of our social and physical world. Before we claim that machines cannot possess *our* contexts for understanding, we need to examine our own. Computers don’t lack context in understanding data. They lack *our* context. We must fight this self-centered notion that our context is what will define a successful form of artificial intelligence. We must remember to look at ourselves more carefully and completely before we make assumptions about our superiority and complexity, and at the same time we must realize that our creations are much younger than we are, and for now refrain from comparing them to ourselves on levels far beyond their physical reach. Instead, we should focus on creating the complex yet small underlying physical structures which allow for such complex results as those in our brain, and build on this foundation. Without a low-level foundation, we are left not with artificial intelligence, but with merely a philosophy.

Bibliography

- Le Cun, Y. et. al. "Handwritten Digit Recognition with a Back-Propagation Network." Updated 1990.
<<http://citeseer.ist.psu.edu/cache/papers/cs/17983/http:zSzzSzwww.research.att.c omzSz~yannzSzexdbzSzpubliszSz.zSzpsgzSzlecun90c.pdf/lecun90handwritten.pdf/>>. Cited 29 April 2004.
- Murray, Bruce. "Understanding brain development and early learning." Updated Date Unspecified. <http://www.facsnet.org/tools/sci_tech/biotek/eliot.php>. Cited 29 April 2004.
- Ndabahaliye, Anicia. "Number of Neurons in a Human Brain." Ed. Glenn Elert. Updated 2002. <<http://hypertextbook.com/facts/2002/AniciaNdabahaliye2.shtml>>. Cited 12 April 2004.
- Pfeffer, Avi. "CS181 Lecture 6 – Decision Trees." Updated 25 February 2004. <<http://www.courses.fas.harvard.edu/~cs181/lectures/lec6.ps>>. Cited 12 April 2004.
- Pfeffer, Avi. "CS181 Lecture 10 – Perceptrons." Updated 10 March 2004. <<http://www.courses.fas.harvard.edu/~cs181/lectures/lec10.ps>>. Cited 12 April 2004.
- Pfeffer, Avi. "CS181 Lecture 12 – Neural Network Model Selection." Updated 17 March 2004. <<http://www.courses.fas.harvard.edu/~cs181/lectures/lec12.ps>>. Cited 12 April 2004.
- Pfeffer, Avi. "CS181 Lecture 17 – Clustering Algorithms." Updated 13 April 2004. <<http://www.courses.fas.harvard.edu/~cs181/lectures/lec17.ps>>. Cited 13 April 2004.
- Zaft, Gordon C. Untitled. Updated 1997. <<http://www.zaft.org/gordon/engr696a/dreyfus.htm>>. Cited 12 April 2004.