Chapter Revision History

The table notes major changes between revisions. Minor changes such as small clarifications or formatting changes are not noted.

Version	Date	Changes	Principal Author(s)
0.4		Initial release	J. Hawkins

Biological and Machine Intelligence: Introduction

The 21st century is a watershed in human evolution. We are solving the mystery of how the brain works and starting to build machines that work on the same principles as the brain. We see this time as the beginning of the era of machine intelligence, which will enable an explosion of beneficial applications and scientific advances.

Most people intuitively see the value in understanding how the human brain works. It is easy to see how brain theory could lead to the cure and prevention of mental disease or how it could lead to better methods for educating our children. These practical benefits justify the substantial efforts underway to reverse engineer the brain. However, the benefits go beyond the near-term and the practical. The human brain defines our species. In most aspects we are an unremarkable species, but our brain is unique. The large size of our brain, and its unique design, is the reason humans are the most successful species on our planet. Indeed, the human brain is the only thing we know of in the universe that can create and share knowledge. Our brains are capable of discovering the past, foreseeing the future, and unravelling the mysteries of the present. Therefore, if we want to understand who we are, if we want to expand our knowledge of the universe, and if we want to explore new frontiers, we need to have a clear understanding of how we know, how we learn, and how to build intelligent machines to help us acquire more knowledge. The ultimate promise of brain theory and machine intelligence is the acquisition and dissemination of new knowledge. Along the way there will be innumerous benefits to society. The beneficial impact of machine intelligence in our daily lives will equal and ultimately exceed that of programmable computers.

But exactly how will intelligent machines work and what will they do? If you suggest to a lay person that the way to build intelligent machines is to first understand how the human brain works and then build machines that work on the same principles as the brain, they will typically say, “That makes sense”. However, if you suggest this same path to artificial intelligence ("AI") and machine learning scientists, many will disagree. The most common rejoinder you hear is “airplanes don’t flap their wings”, suggesting that it doesn’t matter how brains work, or worse, that studying the brain will lead you down the wrong path, like building a plane that flaps its wings.

This analogy is both misleading and a misunderstanding of history. The Wright brothers and other successful pioneers of aviation understood the difference between the principles of flight and the need for propulsion. Bird wings and airplane wings work on the same aerodynamic principles, and those principles had to be understood before the Wright brothers could build an airplane. Indeed, they studied how birds glided and tested wing shapes in wind tunnels to learn the principles of lift. Wing flapping is different; it is a means of propulsion, and the specific method used for propulsion is less important when it comes to building flying machines. In an analogous fashion, we need to understand the principles of intelligence before we can build intelligent machines. Given that the only examples we have of intelligent systems are brains, and the principles of intelligence are not obvious, we must study brains to learn from them. However, like airplanes and birds, we don’t need to do everything the brain does, nor do we need to implement the principles of intelligence in the same way as the brain. We have a vast array of resources in software and silicon to create intelligent machines in novel and exciting ways. The goal of building intelligent machines is not to replicate human behavior, nor to build a brain, nor to create machines to do what humans do. The goal of building intelligent machines is to create machines that work on the same principles as the brain—machines that are able to learn, discover, and adapt in ways that computers can’t and brains can.

Consequently, the machine intelligence principles we describe in this book are derived from studying the brain. We use neuroscience terms to describe most of the principles, and we describe how these principles are

implemented in the brain. The principles of intelligence can be understood by themselves, without referencing the brain, but for the foreseeable future it is easiest to understand these principles in the context of the brain because the brain continues to offer suggestions and constraints on the solutions to many open issues.

This approach to machine intelligence is different than that taken by classic AI and artificial neural networks. AI technologists attempt to build intelligent machines by encoding rules and knowledge in software and human-designed data structures. This AI approach has had many successes solving specific problems but has not offered a generalized approach to machine intelligence and, for the most part, has not addressed the question of how machines can learn. Artificial neural networks (ANNs) are learning systems built using networks of simple processing elements. In recent years ANNs, often called “deep learning networks”, have succeeded in solving many classification problems. However, despite the word “neural”, most ANNs are based on neuron models and network architectures that are incompatible with real biological tissue. More importantly, ANNs, by deviating from known brain principles, don't provide an obvious path to building truly intelligent machines.

Classic AI and ANNs generally are designed to solve specific types of problems rather than proposing a general theory of intelligence. In contrast, we know that brains use common principles for vision, hearing, touch, language, and behavior. This remarkable fact was first proposed in 1979 by Vernon Mountcastle. He said there is nothing visual about visual cortex and nothing auditory about auditory cortex. Every region of the neocortex performs the same basic operations. What makes the visual cortex visual is that it receives input from the eyes; what makes the auditory cortex auditory is that it receives input from the ears. From decades of neuroscience research, we now know this remarkable conjecture is true. Some of the consequences of this discovery are surprising. For example, neuroanatomy tells us that every region of the neocortex has both sensory and motor functions. Therefore, vision, hearing, and touch are integrated sensory-motor senses; we can’t build systems that see and hear like humans do without incorporating movement of the eyes, body, and limbs.

The discovery that the neocortex uses common algorithms for everything it does is both elegant and fortuitous. It tells us that to understand how the neocortex works, we must seek solutions that are universal in that they apply to every sensory modality and capability of the neocortex. To think of vision as a “vision problem” is misleading. Instead we should think about vision as a “sensory motor problem” and ask how vision is the same as hearing, touch or language. Once we understand the common cortical principles, we can apply them to any sensory and behavioral systems, even those that have no biological counterpart. The theory and methods described in this book were derived with this idea in mind. Whether we build a system that sees using light or a system that “sees” using radar or a system that directly senses GPS coordinates, the underlying learning methods and algorithms will be the same.

Today we understand enough about how the neocortex works that we can build practical systems that solve valuable problems today. Of course, there are still many things we don’t understand about the brain and the neocortex. It is important to define our ignorance as clearly as possible so we have a roadmap of what needs to be done. This book reflects the state of our partial knowledge. The table of contents lists all the topics we anticipate we need to understand, but only some chapters have been written. Despite the many topics we don’t understand, we are confident that we have made enough progress in understanding some of the core principles of intelligence and how the brain works that the field of machine intelligence can move forward more rapidly than in the past. The field of machine intelligence is poised to make rapid progress.

Hierarchical Temporal Memory

Hierarchical Temporal Memory, or HTM, is the name we use to describe the overall theory of how the neocortex functions. It also is the name we use to describe the technology used in machines that work on neocortical principles. HTM is therefore a theoretical framework for both biological and machine intelligence.

The term HTM incorporates three prominent features of the neocortex. First, it is best to think of the neocortex as a "memory" system. The neocortex must learn the structure of the world from the sensory patterns that stream into the brain. Each neuron learns by forming connections, and each region of the neocortex is best understood as a type of memory system. Second, the memory in the neocortex is primarily a memory of time-changing, or "temporal", patterns. The inputs and outputs of the neocortex are constantly in motion, usually changing completely several times a second. Each region of the neocortex learns a time-based model of its inputs, it learns to predict the changing input stream, and it learns to play back sequences of motor commands. And finally, the regions of the neocortex are connected in a logical "hierarchy". Because all the regions of the neocortex perform the same basic memory operations, the detailed understanding of one neocortical region leads us to understand how the rest of the neocortex works. These three principles,

"hierarchy", "temporal" patterns, and "memory", are not the only essential principles of an intelligent system, but they suffice as a moniker to represent the overall approach.

Although HTM is a biologically constrained theory, and is perhaps the most biologically realistic theory of how the neocortex works, it does not attempt to include all biological details. For example, the biological neocortex exhibits several types of rhythmic behavior in the firing of ensembles of neurons. There is no doubt that these rhythms are essential for biological brains. But HTM theory does not include these rhythms because we don't believe they play an information-theoretic role. Our best guess is that these rhythms are needed in biological brains to synchronize action potentials, but we don't have this issue in software and hardware implementations of HTM. If in the future we find that rhythms are essential for intelligence, and not just biological brains, then we would modify HTM theory to include them. There are many biological details that similarly are not part of HTM theory. Every feature included in HTM is there because we have an information-theoretical need that is met by that feature.

HTM also is not a theory of an entire brain; it only covers the neocortex and its interactions with some closely related structures such as the thalamus and hippocampus. The neocortex is where most of what we think of as intelligence resides but it is not in charge of emotions, homeostasis, and basic behaviors. Other, evolutionarily older, parts of the brain perform these functions. These older parts of the brain have been under evolutionary pressure for much longer time, and although they consist of neurons, they are heterogeneous in architecture and function. We are not interested in emulating entire brains or in making machines that are human-like, with human-like emotions and desires. Therefore intelligent machines, as we define them, are not likely to pass the Turing test or be like the humanoid robots seen in science fiction. This distinction does not suggest that intelligent machines will be of limited utility. Many will be simple, tirelessly sifting through vast amounts of data looking for unusual patterns. Others will be fantastically fast and smart, able to explore domains that humans are not well suited for. The variety we will see in intelligent machines will be similar to the variety we see in programmable computers. Some computers are tiny and embedded in cars and appliances, and others occupy entire buildings or are distributed across continents. Intelligent machines will have a similar diversity of size, speed, and applications, but instead of being programmed they will learn.

HTM theory cannot be expressed succinctly in one or a few mathematical equations. HTM is a set of principles that work together to produce perception and behavior. In this regard, HTMs are like computers. Computers can't be described purely mathematically. We can understand how they work, we can simulate them, and subsets of computer science can be described in formal mathematics, but ultimately we have to build them and test them empirically to characterize their performance. Similarly, some parts of HTM theory can be analyzed mathematically. For example, the chapter in this book on sparse distributed representations is mostly about the mathematical properties of sparse representations. But other parts of the HTM theory are less amenable to formalism. If you are looking for a succinct mathematical expression of intelligence, you won't find it. In this way, brain theory is more like genetic theory and less like physics.

What is Intelligence?

Historically, intelligence has been defined in behavioral terms. For example, if a system can play chess, or drive a car, or answer questions from a human, then it is exhibiting intelligence. The Turing Test is the most famous example of this line of thinking. We believe this approach to defining intelligence fails on two accounts. First, there are many examples of intelligence in the biological world that differ from human intelligence and would fail most behavioral tests. For example, dolphins, monkeys, and humans are all intelligent, yet only one of these species can play chess or drive a car. Similarly, intelligent machines will span a range of capabilities from mouse-like to super-human and, more importantly, we will apply intelligent machines to problems that have no counterpart in the biological world. Focusing on human-like performance is limiting.

The second reason we reject behavior-based definitions of intelligence is that they don't capture the incredible flexibility of the neocortex. The neocortex uses the same algorithms for all that it does, giving it flexibility that has enabled humans to be so successful. Humans can learn to perform a vast number of tasks that have no evolutionary precedent because our brains use learning algorithms that can be applied to almost any task. The way the neocortex sees is the same as the way it hears or feels. In humans, this universal method creates language, science, engineering, and art. When we define intelligence as solving specific tasks, such as playing chess, we tend to create solutions that also are specific. The program that can win a chess game cannot learn to drive. It is the flexibility of biological intelligence that we need to understand and embed in our intelligent machines, not the ability to solve a particular task. Another benefit of focusing on flexibility is network effects. The neocortex may not always be best at solving any particular problem, but it is very good at solving a huge

array of problems. Software engineers, hardware engineers, and application engineers naturally gravitate towards the most universal solutions. As more investment is focused on universal solutions, they will advance faster and get better relative to other more dedicated methods. Network effects have fostered adoption many times in the technology world; this dynamic will unfold in the field of machine intelligence, too.

Therefore we define the intelligence of a system by the degree to which it exhibits flexibility: flexibility in learning and flexibility in behavior. Since the neocortex is the most flexible learning system we know of, we measure the intelligence of a system by how many of the neocortical principles that system includes. This book is an attempt to enumerate and understand these neocortical principles. Any system that includes all the principles we cover in this book will exhibit cortical-like flexibility, and therefore cortical-like intelligence. By making systems larger or smaller and by applying them to different sensors and embodiments, we can create intelligent machines of incredible variety. Many of these systems will be much smaller than a human neocortex and some will be much larger in terms of memory size, but they will all be intelligent.

About this Book

The structure of this book may be different from those you have read in the past. First, it is a "living book". We are releasing chapters as they are written, covering first the aspects of the theory that are best understood. Some chapters may be published in draft form, whereas others will be more polished. For the foreseeable future this book will be a work in progress. We have a table of contents for the entire book, but even this will change as research progresses.

Second, the book is intended for a technical but diverse audience. Neuroscientists should find the book helpful as it provides a theoretical framework to interpret many biological details and guide experiments. Computer scientists can use the material in the book to develop machine intelligence hardware, software, and applications based on neuroscience principles. Anyone with a deep interest in how brains work or machine intelligence will hopefully find the book to be the best source for these topics. Finally, we hope that academics and students will find this material to be a comprehensive introduction to an emerging and important field that offers opportunities for future research and study.

The structure of the chapters in this book varies depending on the topic. Some chapters are overview in nature. Some chapters include mathematical formulations and problem sets to exercise the reader's knowledge. Some chapters include pseudo-code. Key citations will be noted, but we do not attempt to have a comprehensive set of citations to all work done in the field. As such, we gratefully acknowledge the many pioneers whose work we have built upon who are not explicitly mentioned.

We are now ready to jump into the details of biological and machine intelligence.