Cognitive Computing and Big Data Analytics
By Judith Hurwitz, Marcia Kaufman and Adrian Bowles
Copyright © 2015 by John Wiley & Sons, Inc.

CHAPTER 1

The Foundation of Cognitive Computing

Cognitive computing is a technology approach that enables humans to collaborate with machines. If you look at cognitive computing as an analog to the human brain, you need to analyze *in context* all types of data, from structured data in databases to unstructured data in text, images, voice, sensors, and video. These are machines that operate at a different level than traditional IT systems because they analyze and learn from this data. A cognitive system has three fundamental principles as described below:

- **Learn**—A cognitive system learns. The system leverages data to make inferences about a domain, a topic, a person, or an issue based on training and observations from all varieties, volumes, and velocity of data.
- **Model**—To learn, the system needs to create a model or representation of a domain (which includes internal and potentially external data) and assumptions that dictate what learning algorithms are used. Understanding the context of how the data fits into the model is key to a cognitive system.
- **Generate hypotheses**—A cognitive system assumes that there is not a single correct answer. The most appropriate answer is based on the data itself. Therefore, a cognitive system is probabilistic. A hypothesis is a candidate explanation for some of the data already understood. A cognitive system uses the data to train, test, or score a hypothesis.

This chapter explores the foundations of what makes a system cognitive and how this approach is beginning to change how you can use data to create systems that learn. You can then use this approach to create solutions that change as more data is added (ingested) and as the system learns. To understand how far we have come, you need to understand the evolution of the foundational technologies. Therefore, this chapter provides background information on how artificial intelligence, cognitive science, and computer science have led to the development of cognitive computing. Finally, an overview is provided of the elements of a cognitive computing system.

Cognitive Computing as a New Generation

Cognitive computing is an evolution of technology that attempts to make sense of a complex world that is drowning in data in all forms and shapes. You are entering a new era in computing that will transform the way humans collaborate with machines to gain actionable insights. It is clear that technological innovations have transformed industries and the way individuals conduct their daily lives for decades. In the 1950s, transactional and operational processing applications introduced huge efficiencies into business and government operations. Organizations standardized business processes and managed business data more efficiently and accurately than with manual methods. However, as the volume and diversity of data has increased exponentially, many organizations cannot turn that data into actionable knowledge. The amount of new information an individual needs to understand or analyze to make good decisions is overwhelming. The next generation of solutions combines some traditional technology techniques with innovations so that organizations can solve vexing problems. Cognitive computing is in its early stages of maturation. Over time, the techniques that are discussed in this book will be infused into most systems in future years. The focus of this book is this new approach to computing that can create systems that augment problem-solving capabilities.

The Uses of Cognitive Systems

Cognitive systems are still in the early days of evolution. Over the coming decade you will see cognitive capabilities built into many different applications and systems. There will be new uses that emerge that are either focused on horizontal issues (such as security) or industry-specific problems (such as determining the best way to anticipate retail customer requirements and increase sales, or to diagnose an illness). Today, the initial use cases include some new frontiers and some problems that have confounded industries for decades. For example, systems are being developed that can enable a city manager to anticipate when traffic will be disrupted by weather events and reroute that traffic to avoid problems. In the healthcare industry, cognitive systems are under development that can be used in collaboration with a hospital's electronic medical records to test for omissions and improve accuracy. The cognitive system can help to teach new physicians medical best practices and improve clinical decision making. Cognitive systems can help with the transfer of knowledge and best practices in other industries as well. In these use cases, a cognitive system is designed to build a dialog between human and machine so that best practices are learned by the system as opposed to being programmed as a set of rules.

The list of potential uses of a cognitive computing approach will continue to grow over time. The initial frontier in cognitive computing development has been in the area of healthcare because it is rich in text-based data sources. In addition, successful patient outcomes are often dependent on care providers having a complete, accurate, up-to-date understanding of patient problems. If medical cognitive applications can be developed that enable physicians and caregivers to better understand treatment options through continuous learning, the ability to treat patients could be dramatically improved. Many other industries are testing and developing cognitive applications as well. For example, bringing together unstructured and semi-structured data that can be used within metropolitan areas can greatly increase our understanding of how to improve the delivery of services to citizens. "Smarter city" applications enable managers to plan the next best action to control pollution, improve the traffic flow, and help fight crime. Even traditional customer care and help desk applications can be dramatically improved if systems can learn and help provide fast resolution of customer problems.

What Makes a System Cognitive?

Three important concepts help make a system cognitive: contextual insight from the model, hypothesis generation (a proposed explanation of a phenomenon), and continuous learning from data across time. In practice, cognitive computing enables the examination of a wide variety of diverse types of data and the interpretation of that data to provide insights and recommend actions. The essence of cognitive computing is the acquisition and analysis of the right amount of information in context with the problem being addressed. A cognitive system must be aware of the context that supports the data to deliver value. When that data is acquired, curated, and analyzed, the cognitive system must identify and remember patterns and associations in the data. This iterative process enables the system to learn and deepen its scope so that understanding of the data improves over time. One of the most important practical characteristics of a cognitive system is the capability to provide the knowledge seeker with a series of alternative answers along with an explanation of the rationale or evidence supporting each answer.

A cognitive computing system consists of tools and techniques, including Big Data and analytics, machine learning, Internet of Things (IoT), Natural Language Processing (NLP), causal induction, probabilistic reasoning, and data visualization. Cognitive systems have the capability to learn, remember, provoke, analyze, and resolve in a manner that is contextually relevant to the organization or to the individual user. The solutions to highly complex problems require the assimilation of all sorts of data and knowledge that is available from a variety of structured, semi-structured, and unstructured sources including, but not limited to, journal articles, industry data, images, sensor data, and structured data from operational and transactional databases. How does a cognitive system leverage this data? As you see later in this chapter, these cognitive systems employ sophisticated continuous learning techniques to understand and organize information.

DISTINGUISHING FEATURES OF A COGNITIVE SYSTEM

Although there are many different approaches to the way cognitive systems will be designed, there are some characteristics that cognitive systems have in common. They include the capability to:

- **Learn from experience with data/evidence and improve its own knowledge and performance without reprogramming.**
- **Generate and/or evaluate conflicting hypotheses based on the current state of its knowledge.**
- **Report on findings in a way that justifies conclusions based on confidence in the evidence.**
- **Discover patterns in data, with or without explicit guidance from a user regarding the nature of the pattern.**
- **Emulate processes or structures found in natural learning systems (that is, memory management, knowledge organization processes, or modeling the neurosynaptic brain structures and processes).**
- **Use NLP to extract meaning from textual data and use deep learning tools to extract features from images, video, voice, and sensors.**
- **Use a variety of predictive analytics algorithms and statistical techniques.**

Gaining Insights from Data

For a cognitive system to be relevant and useful, it must continuously learn and adapt as new information is ingested and interpreted. To gain insight and understanding of this information requires that a variety of tools understand the data no matter what the form of the data may be. Today, much of the data required is text-based. *Natural Language Processing* (*NLP*) techniques are needed to capture the meaning of unstructured text from documents or communications from the user. NLP is the primary tool to interpret text. Deep learning tools are required to capture meaning from nontext-based sources such as videos and sensor data. For example, time series analysis analyzes sensor data, whereas a variety of image analysis tools interpret images and videos. All these various types of data have to be transformed so that they can be understood and processed by a machine. In a cognitive system these transformations must be presented in a way that allows the users to understand the relationships between a variety of data sources. Visualization tools and techniques will be critical ways for making this type of complex data accessible and understandable. *Visualization* is one of the most powerful techniques to make it easier to recognize patterns in massive and complex data. As we evolve to cognitive computing we may be required to bring together structured, semi-structured, and unstructured sources to continuously learn and gain insights from data. How these data sources are combined with processes for gaining results is key to cognitive computing. Therefore, the cognitive system offers its users a different experience in the way it interacts with data and processes.

Domains Where Cognitive Computing Is Well Suited

Cognitive computing systems are often used in domains in which a single query or set of data may result in a hypothesis that yields more than one possible answer. Sometimes, the answers are not mutually exclusive (for example, multiple, related medical diagnoses where the patient may have one or more of the indicated disorders at the same time). This type of system is probabilistic, rather than deterministic. In a *probabilistic system*, there may be a variety of answers, depending on circumstances or context and the confidence level or probability based on the system's current knowledge. A *deterministic system* would have to return a single answer based on the evidence, or no answer if there were a condition of uncertainty.

The cognitive solution is best suited to help when the domain is complex and conclusions depend on who is asking the question and the complexity of the data. Even though human experts might know an answer to a problem, they may not be aware of new data or new circumstances that will change the outcome of an inquiry. More advanced systems can identify missing data that would change the confidence level of an answer and request further information interactively to converge on an answer or set of answers with sufficient confidence to help the user take some action. For example, in the medical diagnostic example, the cognitive system may ask the physician to perform additional tests to rule out or to choose certain diagnoses.

DEFINING NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is the capability of computer systems to process text written or recorded in a language used for human communication (such as English or French). Human "natural language" is filled with ambiguities. For example, one word can have multiple meanings depending on how it is used in a sentence. In addition, the meaning of a sentence can change dramatically just by adding or removing a single word. NLP enables computer systems to interpret the meaning of language and to generate natural language responses.

Cognitive systems typically include a knowledge base (corpus) that has been created by ingesting various structured and unstructured data sources. Many of these data sources are text-based documents. NLP is used to identify the semantics of words, phrases, sentences, paragraphs, and other linguistic units in the documents and other unstructured data found in the corpus. One important use of NLP in cognitive systems is to identify the statistical patterns and provide the linkages in data elements so that the meaning of unstructured data can be interpreted in the right context.

For more information on natural language processing, see Chapter 3, "Natural Language Processing in Support of a Cognitive System."

Artificial Intelligence as the Foundation of Cognitive Computing

Although the seeds of artificial intelligence go back at least 300 years, the evolution over the past 50 years has had the most impact for cognitive computing. Modern *Artificial Intelligence* (*AI*) encompassed the work of scientists and mathematicians determined to translate the workings of neurons in the brain into a set of logical constructs and models that would mimic the workings of the human mind. As computer science evolved, computer scientists assumed that it would be possible to translate complex thinking into binary coding so that machines could be made to think like humans.

Alan Turing, a British mathematician whose work on cryptography was recognized by Winston Churchill as critical to victory in WWII, was also a pioneer in computer science. Turing turned his attention to machine learning in the 1940s. In his paper called "Computing Machinery and Intelligence" (written in 1950 and published in *Mind,* a United Kingdom peer-reviewed academic journal), he posed the question, "Can machines think?" He dismissed the argument that machines could never think because they possess no human emotion. He postulated that this would imply that "the only way to know that a man thinks is to be that particular man. . . ." Turing argued that with advancement in digital computing, it would be possible to have a learning machine whose internal processes were unknown, or a black box. Thus, "its teacher will often be very largely ignorant of quite what is going on inside, although he will still be able to some extent to predict his pupil's behavior."

In his later writing Turing proposed a test to determine if a machine possessed intelligence, or could mimic the behaviors we associate with intelligence. The test consisted of two humans and a third person that inputted questions for the two people via a typewriter. The goal of the game was to determine if the game players could determine which of the three participants was a human and which was a "typewriter" or a computer. In other words, the game consisted of human/machine interactions. It is clear that Turing was ahead of his time. He was making the distinction between the ability of the human to intuitively operate in a complex world and how well a machine can mimic those attributes.

Another important innovator was Norbert Weiner, whose 1948 book, *Cybernetics or Control and Communication in the Animal and the Machine*, defined the field of cybernetics. While working on a World War II research project at MIT, he studied the continuous feedback that occurred between a guided missile system and its environment. Weiner recognized that this process of continuous feedback occurred in many other complex systems including machines, animals, humans, and organizations. Cybernetics is the study of these feedback mechanisms. The feedback principle describes how complex systems (such as the guided missile system) change their actions in response to their environment. Weiner's theories on the relationship between intelligent behavior and feedback mechanisms led him to determine that machines could simulate human feedback mechanisms. His research and theories had a strong influence on the development of the field of AI.

Games, particularly two-person zero-sum perfect information games (in which both parties can see all moves and can theoretically generate and evaluate all future moves before acting), have been used to test ideas about learning behavior since the dawn of AI. Arthur Lee Samuel, a researcher who later went to work for IBM, developed one of the earliest examples. He is credited with developing the first self-learning program for playing checkers. In his paper published in the *IBM Journal of Research and Development* in 1959, Samuel summarized his research as follows:

> *Two machine-learning procedures have been investigated in some detail using the game of checkers. Enough work has been done to verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program. Furthermore, it can learn to do this in a remarkably short period of time (8 or 10 hours of machine-playing time) when given only the rules of the game, a sense of direction, and a redundant and incomplete list of parameters which are thought to have something to do with the game, but whose correct signs and relative weights are unknown and unspecified. The principles of machine learning verified by these experiments are, of course, applicable to many other situations.*

Samuel's research was an important precursor to the work that followed over the coming decades. His goal was not to find a way to beat an opponent in checkers, but to figure out how humans learned. Initially, in Samuel's checkers experiment, the best he achieved was to have the computer play to a draw with the human opponent.

In 1956, researchers held a conference at Dartmouth College in New Hampshire that helped to define the field of AI. The participants included the most important researchers in what was to become the field of AI. The participants included Allen Newell and Herbert A. Simon of Carnegie Tech (Carnegie Mellon University), Marvin Minsky from MIT, and John McCarthy (who left MIT in 1962 to form a new lab at Stanford). In their proposal for the Dartmouth event, McCarthy et al. outlined a fundamental conjecture that influenced AI research for decades: "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." (McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955), "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.") Also in 1956, Allen Newell, Herbert Simon, and Cliff Shaw created a program called the "Logic Theorist" that is possibly the first AI computer program. It was created to prove mathematical theorems by simulating certain human problem-solving capabilities.

Herbert Simon, who won the Nobel Prize for Economics in 1978, had an ongoing interest in human cognition and decision making that factored into all his research. He theorized that people are rational agents who can adapt to conditions. He assumed that there could be a simple interface between human knowledge and an artificially intelligent system. Like his predecessors, he assumed that it would be relatively easy to find a way to represent knowledge as an information system. He contended that transition to AI could be accomplished by simply adapting rules based on changing requirements. Simon and his colleagues such as Alan Newell assumed that a simple adaptive mechanism would allow intelligence to be captured to create an intelligent machine.

One of Simon's important contributions to the burgeoning field was an article he wrote about the foundational elements and the future of capturing intelligence. Simon laid out the concept of natural language processing and the capability of computers to mimic vision. He predicted that computers would play chess at the grand master level. (Allen Newell, Cliff Shaw, Herbert Simon. "Chess Playing Programs and the Problem of Complexity." *IBM Journal of Research and Development*, Vol. 4, No. 2, 1958.)

Although many of the early endeavors were wildly optimistic, they did send the field of AI in the right direction. Many of the computer scientists assumed that within 20 years computers would be capable of mimicking cognitive processes fundamental to learning. When many commercial AI start-ups failed to create ongoing businesses in the 1980s, it became clear that new research and more time were needed to fulfill expectations for commercial applications in the field of AI. Scientists and researchers continued to innovate in areas such as symbolic reasoning, expert systems, pattern recognition, and machine learning. In addition, there were extensive developments in related and parallel areas such as robotics and neural networks.

Another significant contributor to AI research was Professor Edward Feigenbaum. In 1965, after joining the computer science faculty at Stanford University, Feigenbaum and Nobel laureate Joshua Lederberg started the DENDRAL project, which was later referred to as the first expert system. The project's importance to the field of AI is based on the framework that was created for other expert systems to follow. Feigenbaum said that the DENDRAL project was important because it showed that "the dream of a really intelligent machine was possible. There was a program that was performing at world class levels of problem-solving competence on problems that only Ph.Ds. solve—these mass spectral analysis problems." Today, expert systems are used in the military and in industries such as manufacturing and healthcare.

EXPERT SYSTEMS DEFINED

Expert system technology has been around for decades and gained popularity in the 1980s. An *expert system* captures knowledge from domain experts in a knowledge or rules base. The developer of an expert system needs to determine the rules in advance. Occasionally, there are confidence factors applied to the data in the knowledge base. When changes occur, the expert system needs to be updated by a subject matter expert. An expert system is most useful when there is an area of knowledge that will not change dramatically over time. After data is ingested into the system, it can be used to assess different hypotheses and determine the consequences of an assertion. In addition, an expert system can use fuzzy logic as a way to assess the probability of a specific rule included within the system. Often, expert systems are used as a classification technique to help determine how to manage unstructured data.

The U.S. Defense Advanced Research Projects Agency (DARPA) funded much of the underlying research in AI. The agency is responsible for the development of new technologies that can be used by the military. Prior to 1969, millions of dollars were provided for AI research with limited or no direction as to the type of research activities. However, after 1969, DARPA funding was legally restricted to be applied to specific military projects such as autonomous tanks and battle management systems. Expert systems were designed to provide guidance to personnel in the field. Many of these AI systems codified best practices by studying historical events. For example, in the late 1980s DARPA sponsored the FORCES project, which was part of the Air Land Battle Management Program. This was an expert system designed to help field personnel make decisions based on historical best practices. A commander using the system could ask, "What would General Patton do now?" This system was not actually deployed, but provided good experience for knowledge-based defense projects that were built later.

During the 1970s and 1980s there were significant periods of time when it became difficult for scientists to receive funding for AI projects. Although military-based research continued to be funded by DARPA, commercial based funding was almost non-existent. In some cases, computer scientists looking for grants for research would use the term "expert systems" or "knowledge-based systems" rather than AI to help ensure funding. However, subfields of AI including machine learning, ontologies, rules management, pattern matching, and NLP continued to find their way into a myriad of products over the years. Even the Automated Teller Machine (ATM) has evolved to incorporate many of these technologies.

One of the early commercial projects that took AI and machine learning back into prominence was a project initiated by American Express. The project was designed to look for patterns of fraud in credit card transactions. The results of the project were wildly successful. Suddenly, a technology approach that had been maligned was showing business value. The secret ingredient that made the project work was that American Express fed this system a massive amount of data. Typically, companies found it much too expensive to store this much data. American Express gambled that the investment would be worth the price. The results were dramatic. By detecting patterns that would result in fraud, American Express saved an enormous amount of money. The American Express project leveraged machine learning combined with huge volumes of data to determine that fraud was about to take place and stopped those transactions. This was one of the early indications that machine learning and pattern-based algorithms could become an engine for business transformation. It was the beginning of the reinvestment in the emerging field of machine learning—a field that took its foundation from the concepts of AI.

AI is focused on determining how to represent knowledge in a way that the data can be manipulated so that people can make inferences from that knowledge. The field has evolved over the decades. Today, most of the focus is on the area of machine learning algorithms that provide a mechanism to allow computers to process data in a methodical way. But much of the focus of machine learning is dealing with ambiguity because most data is unstructured and open to many different interpretations.

Understanding Cognition

Understanding how the human brain works and processes information provides a blueprint for the approach to cognitive computing. However, it is not necessary to build a system that replicates all the capabilities of the human brain to serve as a good collaborator for humans. By understanding cognition we can build systems that have many of the characteristics required to continuously learn and adapt to new information. The word *cognition*, from the Latin root gnosis, meaning to know and learn, dates back to the 15th century. Greek philosophers were keenly interested in the field of deductive reasoning.

With cognitive computing, we are bringing together two disciplines:

- **Cognitive science**—The science of the mind.
- **Computer science**—The scientific and practical approach to computation and its applications. It is the systematic technique for translating this theory into practice.

The main branches of cognitive science are psychology (primarily an applied science, in helping diagnose and treat mental/behavioral conditions) and neurology (also primarily applied, in diagnosis/treatment of neurological conditions). Over the years, however, it became clear that there was a critical relationship between the way the human brain works and computer engineering. For example, cognitive scientists, in studying the human mind, have come to understand that human cognition is an interlinking system of systems that allows for information to be received from outside inputs, which is then stored, retrieved, transformed, and transmitted. Likewise, the maturation of the computer field has accelerated the field of cognitive sciences. Increasingly, there is less separation between these two disciplines.

A foundational principle of cognitive science is that an intelligent system consists of a number of specialized processes and services (within the human brain) that interact with each other. For example, a sound transmits a signal to the brain and causes a person to react. If the loud sound results in pain, the brain learns to react by causing the human to place her hands over her ears or by moving away. This isn't an innate reaction; it is learned as a response to a stimulus. There are, of course, different variations in cognition, depending on differences in genetic variations. (A deaf person reacts differently to sound than a person who hears well.) However, these variations are the exception, not the rule.

To make sense of how different processes in the brain relate to each other and impact each other, cognitive scientists model cognitive structures and processes. There isn't a single cognitive architecture; rather, there are many different approaches, depending on the interaction model. For example, there may be an architecture that is related to human senses such as seeing, understanding speech, and reacting to tastes, smells, and touch. A cognitive architecture is also directly tied to how the neurons in the brain carry out specific tasks, absorb new inputs dynamically, and understand context. All this is possible even if there is sparse data because the brain can fill in the implied information. The human brain is architected to deal with the mental processes of perception, memory, judgment, and learning. Humans can think fast and draw conclusions based on their ability to reason or make inferences from the pieces of information they are given.

Humans have the ability to make speculative conjectures, construct imaginative scenarios, use intuition, and other cognitive processes that go beyond mere reasoning, inference, and information processing. The fact that humans have the ability to come up with a supposition based on sparse data points to the brilliance of human cognition. However, there can be negative consequences of this inference. The human may have a bias that leads to conclusions that are erroneous. For example, the human may look at one research study that states that there are some medical benefits to chocolate and conclude that eating a lot of candy will be a good thing. In contrast, a cognitive architecture will not make the mistake of assuming that one study or one conclusion has an overwhelming relevance unless there is actual evidence to draw conclusions. Unlike humans, machines do not have bias unless that bias is programmed into the system.

Traditional architectures rely on humans to interpret processes into code. AI assumes that computers can replace the thinking process of humans. With cognitive computing, the human leverages the unique ability of computers to process, manage, and associate information to expand what is possible.

Two Systems of Judgment and Choice

It is quite complicated to translate the complexity of human thought and actions into systems. In human systems, we are often influenced by emotion, instinct, habits, and subconscious assumptions about the world. *Cognition* is a foundational approach that leverages not just how we think, but also how we act and how we make decisions. Why does one doctor recommend one treatment whereas another doctor recommends a completely different approach to the same disease? Why do two people raised in the same household with a similar experience grow up to have diametrically opposed views of the world? What explains how we come to conclusions and what does this tell us about cognition and cognitive computing?

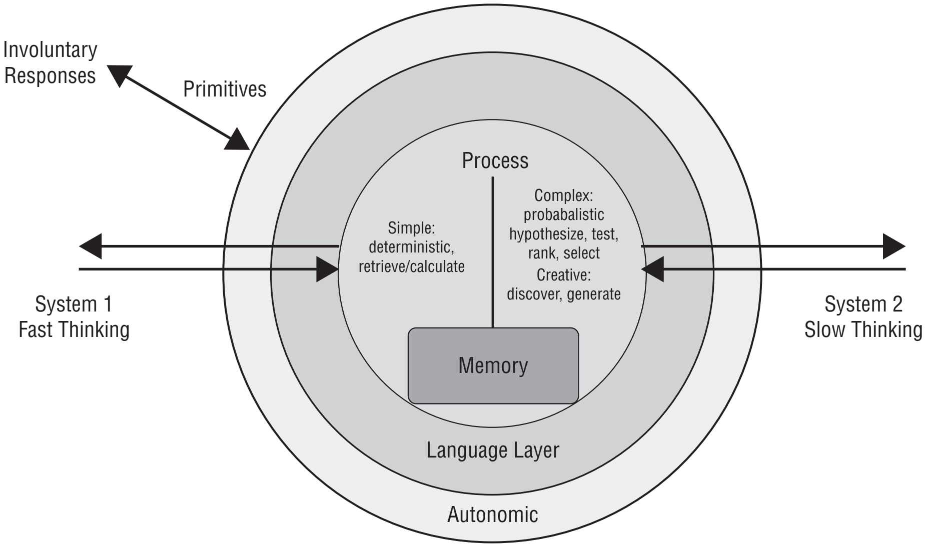
One of the most influential thinkers on the topic is Dr. Daniel Kahneman, an Israeli-American psychologist and winner of the 2002 Nobel Memorial Prize in Economic Sciences. He is well known for his research and writing in the field of the psychology of judgment and decision making. One of his greatest contributions to cognitive computing is his research on the cognitive basis for common human errors that arise from heuristic and biases. To understand how to apply cognition to computer science, it is helpful to understand Kahneman's theory about how we think. In 2011, he published a book, *Thinking Fast and Slow*, which provides important insights for cognitive computing. The following section provides some insights into Kahneman's thinking and how it relates to cognitive computing. Kahneman divides his approach to judgment and reasoning into two forms: System 1: Intuitive thinking, and System 2: Controlled and rule-centric thinking.

This next section describes these two systems of thought and how they relate to how cognitive computing works. System 1 thinking is the type of intuitive reasoning that can be analogous to the type of processing that can be easily automated. In contrast, System 2 thinking is the way we process data based on our experiences and input from many data sources. System 2 thinking is related to the complexities of cognitive computing.

System 1—Automatic Thinking: Intuition and Biases

System 1 thinking is what happens automatically in our brains. It uses our intuition to draw conclusions. Therefore, it is relatively effortless. System 1 thinking begins almost from the moment we are born. We learn to see objects and understand their relationships to ourselves. For example, we associate our mother's voice with safety. We associate a loud noise with danger. These associations form the basis of how we experience the world. The child with a cruel mother will not have the same association with the mother's voice as the child with the kind mother. Of course, there are other issues at play as well. The child with a kind mother may have an underlying mental illness that causes irrational actions. An average child who associates a loud noise with fun may not feel in danger. As people learn over time, they begin to assimilate automatic thinking into their way of operating in the world. The chess protégée who becomes a master automatically learns to make the right moves. The chess master not only knows what his next move should be but also can anticipate what move his opponent will do next. That chess master can play an entire game in his mind without even touching the chessboard. Likewise, emotions and attitudes about the world are automatic, as well. If a person is raised in a dangerous area of a city, he will have automatic attitudes about those people around him. Those attitudes are not something that he even thinks about and cannot easily be controlled. These attitudes are simply part of who he is and how he has assimilated his environment and experiences.

The benefit of System 1 thinking is that we can take in data from the world around us and discover the connections between events. It is easy to see that System 1 is important to cognitive computing because it allows us as humans to use sparse information we collect about events and observations and come to rapid conclusions. System 1 can generate predictions by matching these observations. However, this type of intuitive thinking can also be inaccurate and prone to error if it is not checked and monitored by what Kahneman calls System 2: the ability to analyze massive amounts of information related to the problem being addressed and to reason in a deliberate manner. Combining System 1 intuitive thinking with System 2 deep analysis is critical for cognitive computing. Figure 1-1 shows the interaction between intuitive thinking and deep analysis.


Involuntary Responses
Primitives
Process
Simple: deterministic, retrieve/calculate
Complex: probabalistic hypothesize, test, rank, select
Creative: discover, generate
System 1 Fast Thinking
System 2 Slow Thinking
Memory
Language Layer
Autonomic

Copyright (c) 2014 by STORM insights, Inc. All Rights Reserved.

Figure 1-1: Interaction between intuitive thinking and deep analysis

System 2—Controlled, Rule-Centric, and Concentrated Effort

Unlike System 1 thinking, System 2 thinking is a reasoning system based on a more deliberate process. System 2 thinking observes and tests assumptions and observations, instead of jumping to a conclusion based on what is assumed. System 2 thinking uses simulation to take an assumption and look at the implications of that assumption. This type of system requires that we collect a lot of data and build models that test System 1 intuition. This is especially important because System 1 thinking is typically based on a narrow view of a situation: a silo. Although an idea may appear to be good and plausible when viewed from a narrow perspective, when viewed in context with other data, conclusions often change. Drug trials are an excellent example of this phenomenon. A potential cancer treatment seems promising. All the preliminary data indicates that the drug will eradicate the cancer cells. However, the treatment is so toxic that it also destroys healthy cells. System 1 thinking would assume that the fact that cancer cells are destroyed is enough to determine that the drug should immediately be put on the market. However, System 1 thinking often includes bias. Although it may appear that an approach makes sense, the definition of the problem may be ill-defined. System 2 thinking slows down the evaluation process and looks at the full context of the problem, collects more data across silos, and comes up with a solution. Because System 2 is anchored in data and models, it takes into account those biases and provides a better outcome. Predicting outcomes is a complex business issue because so many factors can change outcomes. This is why it is important to combine intuitive thinking with computational models.

Understanding Complex Relationships Between Systems

Because of the advent of cognitive computing, we are beginning to move beyond the era in which a system must be designed as a unified environment intended to solve a specific, well-defined problem. In this new world, complex systems are not necessarily massive programs. Rather, they may be developed as modular services that execute specific functions and are intended to operate based on the actions and the data from specific events. These adaptive systems are designed so that they can be combined with other elements at the right time to determine the answer to a complex problem. What makes this difficult is the requirement to integrate data from a variety of sources. The process begins with the physical ability to ingest data sources. However, the real complexity is both the integration process and the process of discovering relationships between data sources. Unstructured text-based information sources have to be parsed so that it is clear what content is the proper nouns, verbs, and objects. This process of categorization is necessary so that the data can be consistently managed. Data from unstructured sources such as images, video, and voice have to be analyzed through deep analytics of patterns and outliers. For example, recognition of human facial images may be facilitated by analyzing the edge of the image and identifying for patterns that can be interpreted as objects—such as a nose versus an eye. Analysis is done to get a central category based on evaluating all the data in context. The key to success in this complicated process is to ingest enough data in these categories so that it is possible to apply a machine-learning algorithm that can continue to refine the data. The broader the knowledge area is, the more difficult this process will be.

When data is combined from a variety of sources, it must be categorized into some sort of database structure. It is most helpful to have an approach that is highly interdisciplinary and provides a framework to help individuals find answers to some fundamental questions based on continually refining the elements of the information sources that are most relevant. For example, if the system can decipher a proper noun and then find verbs and the object of that verb, it will be easier to determine the context for the data so that the user can make sense of that data and apply it to a problem domain.

ANALYZING IMAGES, VIDEO, AND AUDIO

The human brain has the ability to automatically translate images into meaning. A doctor who is trained to read an x-ray can interpret differences in results in hundreds of patients in near-real time. The untrained individual can possibly recognize a picture of a person he has only met twice. Being able to extract data from images, videos, and speech is an important issue in gaining understanding of all types of data. This type of analytics has been helped significantly with the advent of cloud-based services. These services make it possible to scale advanced analytics on everything from machine vision, speech recognition, and the ability to gain insights into real-time streaming of images and video. In a cognitive system it is critical to be able to analyze this information to gain insights into information that is not text based. For example, analyzing image data from thousands of faces may identify a criminal or terrorist. Analyzing motion and sound data may provide insights into the severity of an earthquake. Using sophisticated algorithms help determine patterns in this type of unstructured or semi-structured data.

Types of Adaptive Systems

Cognitive systems are intended to address real-world problems in an adaptive manner. This adaptive systems approach is intended to deliver relevant data-driven insights to decision makers based on advanced analysis of the data. The knowledge base is managed and updated as needed to ensure that the full semantic context of the data is leveraged in the analytic process. For example, the system could be looking at the stock market and the complex set of information about individual companies, statistics about performance of economies, and competitive environments. The goal of the adaptive system would be to bring these elements together so that the consumer of that system gains a holistic view of the relationship between factors. An adaptive system approach can be applied to medicine so that a physician can use a combination of learned knowledge and a corpus of knowledge from clinical trials, research, and journal articles to better understand how to treat a disease.

The combination of computer and human interactions enables cognitive systems to gain a dynamic and holistic view of a specific topic or domain. To be practical, many elements have to come together with the right level of context and right amounts of information from the right sources. These elements have to be coordinated based on the principles of self-organization that mimic the way the human brain assimilates information, draws conclusions, and tests those conclusions. This is not simple to execute. It requires that there is enough information from a variety of sources. The system must therefore discover, digest, and adapt a massive amount of data. The system must look for patterns and relationships that aren't visible to the unassisted human. These types of adaptive systems are an attempt to mimic the way the human brain makes associations—often on sparse data.

The Elements of a Cognitive System

A cognitive system consists of many different elements, ranging from the hardware and deployment models to machine learning and applications. Although many different approaches exist for creating a cognitive system, there are some common elements that need to be included. Figure 1-2 shows an overview of the architecture for a cognitive system, which is described in the next section. Chapter 2, “Design Principles for Cognitive Systems,” goes deeper into a discussion of each of these elements.

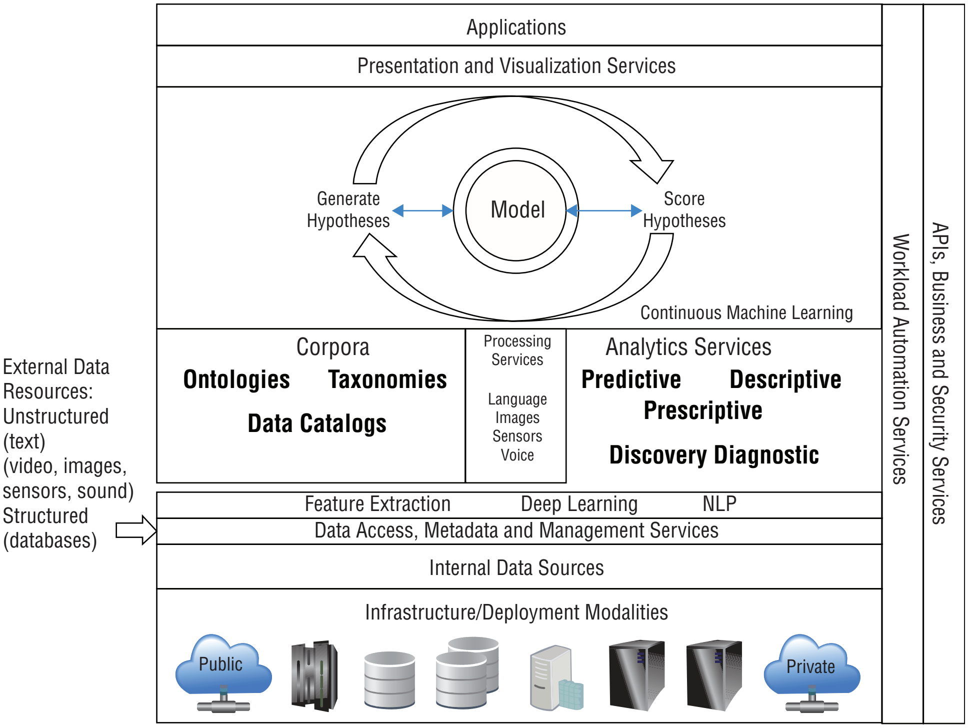

Applications
Presentation and Visualization Services
Generate Hypotheses
Model
Score Hypotheses
Continuous Machine Learning
Corpora
Ontologies
Taxonomies
Data Catalogs
Processing Services
Language
Images
Sensors
Voice
Analytics Services
Predictive
Descriptive
Prescriptive
Discovery Diagnostic
External Data Resources: Unstructured (text) (video, images, sensors, sound) Structured (databases)
Feature Extraction
Deep Learning
NLP
Data Access, Metadata and Management Services
Internal Data Sources
Infrastructure/Deployment Modalities
Public
Private
Workload Automation Services
APIs, Business and Security Services

Figure 1-2: Elements of a cognitive system

Infrastructure and Deployment Modalities

In a cognitive system it is critical to have a flexible and agile infrastructure to support applications that continue to grow over time. As the market for cognitive solutions matures, a variety of public and private data need to be managed and processed. In addition, organizations can leverage Software as a Service (SaaS) applications and services to meet industry-specific requirements. A highly parallelized and distributed environment, including compute and storage cloud services, must be supported.

Data Access, Metadata, and Management Services

Because cognitive computing centers around data, it is not surprising that the sourcing, accessing, and management of data play a central role. Therefore, before adding and using that data, there has to be a range of underlying services. To prepare to use the ingested data requires an understanding of the origins and lineage of that data. Therefore, there needs to be a way to classify the characteristics of that data such as when that text or data source was created and by whom. In a cognitive system these data sources are not static. There will be a variety of internal and external data sources that will be included in the corpus. To make sense of these data sources, there needs to be a set of management services that prepares data to be used within the corpus. Therefore, as in a traditional system, data has to be vetted, cleansed, and monitored for accuracy.

The Corpus, Taxonomies, and Data Catalogs

Tightly linked with the data access and management layer are the corpus and data analytics services. A *corpus* is the knowledge base of ingested data and is used to manage codified knowledge. The data required to establish the domain for the system is included in the corpus. Various forms of data are ingested into the system (refer to Figure 1-2). In many cognitive systems, this data will primarily be text-based (documents, textbooks, patient notes, customer reports, and such). Other cognitive systems include many forms of unstructured and semi-structured data (such as videos, images, sensors, and sounds). In addition, the corpus may include ontologies that define specific entities and their relationships. *Ontologies* are often developed by industry groups to classify industry-specific elements such as standard chemical compounds, machine parts, or medical diseases and treatments. In a cognitive system, it is often necessary to use a subset of an industry-based ontology to include only the data that pertains to the focus of the cognitive system. A taxonomy works hand in hand with ontologies. A *taxonomy* provides context within the ontology.

Data Analytics Services

Data analytics services are the techniques used to gain an understanding of the data ingested and managed within the corpus. Typically, users can take advantage of structured, unstructured, and semi-structured data that has been ingested and begin to use sophisticated algorithms to predict outcomes, discover patterns, or determine next best actions. These services do not live in isolation. They continuously access new data from the data access layer and pull data from the corpus. A number of advanced algorithms are applied to develop the model for the cognitive system.

Continuous Machine Learning

Machine learning is the technique that provides the capability for the data to learn without being explicitly programmed. Cognitive systems are not static. Rather, models are continuously updated based on new data, analysis, and interactions. A machine learning process has two key elements: hypothesis generation and hypothesis evaluation. Machine learning is discussed in detail in Chapter 2.

Hypothesis Generation and Evaluation

A *hypothesis* is a testable assertion based on evidence that explains some observed phenomenon. In a cognitive computing system, you look for evidence to support or refute hypotheses. You need to acquire data from various sources, create models, and then test how well the models work. This is done through an iterative process of training the data. Training may occur automatically based on the systems analysis of data, or training may incorporate human end users. After training, it begins to become clear if the hypothesis is supported by the data. If the hypothesis is not supported by the data, the user has several options. For example, the user may refine the data by adding to the corpus, or change the hypothesis. To evaluate the hypothesis requires a collaborative process of constituents that use the cognitive system. Just as with the creation of the hypothesis, the evaluation of results refines those results and trains again.

The Learning Process

To learn from data you need tools to process both structured and unstructured data. For unstructured textual data, NLP services can interpret and detect patterns to support a cognitive system. Unstructured data such as images, videos, and sound requires deep learning tools. Data from sensors are important in emerging cognitive systems. Industries ranging from transportation to healthcare use sensor data to monitor speed, performance, failure rates, and other metrics and then capture and analyze this data in real time to predict behavior and change outcomes. Chapter 2 discusses the tools used to process the varied forms of data analyzed in a cognitive system.

Presentation and Visualization Services

To interpret complex and often massive amounts of data requires new visualization interfaces. *Data visualization* is the visual representation of data as well as the study of data in a visual way. For example, a bar chart or pie chart is a visual representation of underlying data. Patterns and relationships in data are easier to identify and understand when visualized with structure, color, and such. The two basic types of data visualizations are static and dynamic. In either or both cases, there may also be a requirement for interactivity. Sometimes looking at the visualized representation of the data is not enough. You need to drill down, re-position, expand and contract, and so on. This interactivity enables you to "personalize" the views of the data so that you can pursue non-obvious manifestations of data, relationships, and alternatives. Visualization may depend on color, location, and proximity. Other critical issues that impact visualization include shape, size, and motion. Presentation services prepare results for output. Visualization services help to communicate results by providing a way to demonstrate the relationships between data.

A cognitive system brings text or unstructured data together with visual data to gain insights. In addition, images, motion, and sound are also elements that need to be analyzed and understood. Making this data interactive through a visualization interface can help a cognitive system be more accessible and usable.

Cognitive Applications

A cognitive system must leverage underlying services to create applications that address problems in a specific domain. These applications that are focused on solving specific problems must engage users so that they gain insights and knowledge from the system. In addition, these applications may need to infuse processes to gain insight about a complex area such as preventive maintenance or treatment for a complex disease. An application may be designed to simulate the smartest customer service agent. The end goal is to turn an average employee into the smartest employee with many years of experience. A well-designed cognitive system provides the user with contextual insights based on role, the process, and the customer issue they are solving. The solution should provide the users insights so they make better decisions based on data that exists but is not easily accessible.

Summary

A cognitive computing system is intended to provide a platform to solve hypotheses based on learning from data. These systems are best used to solve problems in data-rich domains. A probabilistic approach to systems design is helping to create a new generation of systems that will focus on helping make sense of a complex world.