Chapter 1
Smart Technologies—Scope and Applications

K. B. Akhilesh

1.1 Introduction

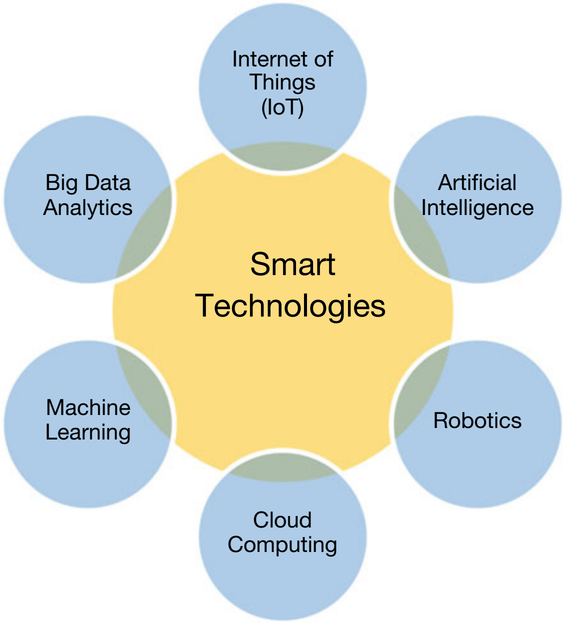
The next-generation technologies that are going to impact our lives in the coming future are being conceptualized, developed and deployed at an increasing rate. Until recently, contemporary technologies were being developed in silos and task-specific applications were being developed in a stand-alone manner. As compared to this, the next-generation technologies are being developed in such a way that they will allow integration of these silo applications under one large umbrella network by improving machine-to-machine connectivity and compatibility. Aggregation of applications under one overarching technology framework has its own set of merits and demerits. Such kind of large-scale aggregation can increase the applicability, enhance the real-time data collection and improve the decision-making ability of users. However, aggregation of such technologies significantly increases the complexity, vulnerability and expenditure required to set up these technology platforms. In this chapter, we will briefly introduce some of these 'smart' technologies namely Internet of things (IoT), cyber-physical systems, artificial intelligence, robotics, cloud computing, machine learning and big data analytics, which are currently being developed to make our lives 'smarter' (Fig. 1.1).

Cutting-edge research, innovation and entrepreneurship is being undertaken at academia, industry and government agencies around these smart technologies and their possible adoption in the daily lives of the people in the form of novel products and services. These smart technologies are slowly evolving in a theme-based manner, where integrated solutions are being designed to solve a particular unmet need of the target audience through converge of diverse technologies. Traditional technologies in hardware, software and networking are being seamlessly integrated to create these 'smart' innovations. For example, conventional Embedded Systems are being

K. B. Akhilesh (✉)
Department of Management Studies, Indian Institute of Science, Bengaluru, India
e-mail: kba@iisc.ac.in

K. B. Akhilesh and D. P. F. Möller (eds.), *Smart Technologies*,
https://doi.org/10.1007/978-981-13-7139-4_1

Fig. 1.1 Smart Technologies



upgraded with multiple sensors, real-time processors and high-speed communication networks to create 'Internet-of-Things' (IoT) platforms. Traditional mechanical systems are being replaced by highly sophisticated robots which can enhance the effectiveness and efficiency of tasks at workplaces such as factories, laboratories and even homes. The use of robots in situations deemed 'risky' for humans is on the rise and increasingly humans are being replaced by robots in such potentially hazardous situations. The exponential rise of computing power has also enabled processors to handle large datasets with ease and affordability leading to the popularity of 'Big Data Analytics'. Decision-makers are increasingly relying on these statistical algorithms to gain a better understanding about the past performance, latest trends and future insights about their topic of interest. Cloud Computing is ensuring that computing resources and data storage space is readily available for clients on-demand on rental basis thus eliminating the need for them to purchase expensive resources. The remaining part of the chapter briefly introduces each of these technologies and explains key aspects associated with them, along with their potential areas of application.

1.2 Internet of Things (IoT)

Internet of things (IoT) is one technology framework which is being expected to transform key areas like security, agriculture, manufacturing, healthcare, transportation, education, retail, insurance, banking and financial services, etc. This framework consists of physical devices embedded with sensors, software and connecting modules which enable them to connect with other devices in the network and share data on a real-time basis. Each device is given a unique address and an ability to communicate with any other device existing in the interconnected network. This framework facilitates seamless real-time integration of the physical devices with the computing systems over the 'cloud' network resulting in a better efficiency, effectiveness and speed of operation by eliminating human intervention. This allows devices and systems around us to become smarter and better integrated with the ambient world around us through regular data collection and analysis. Such integration of physical (hardware) systems along with computer-based algorithms, the worldwide Internet and its users has also been termed as 'cyber-physical systems.'

The first level of integration between the hardware systems and the software systems improves the capability, functionality and flexibility of these basic devices. The next level of integration allows such devices to connect to the worldwide internet and transfer real-time data which further improves its 'connectivity' to the wider world, outside its dedicated application. Such improved functionality and connectivity of previously 'stand-alone' devices improves their 'smartness' quotient which can lead to game-changing possibilities.

One of the major challenges involved in the entire process of designing such 'complex' systems revolves around the convergence of computational elements, cyber-physical systems and multiple networking protocols in a single product and its operating process. The success of such systems will largely depend on the 'seamless' coordination between the computational, physical and networking components that are integral to such systems. Only then such systems can successfully be implemented in areas as diverse as homes, workplaces, transportation, manufacturing, health care, education and entertainment, etc. The other challenge arises with respect to the robustness of the connected device network to malicious attacks. Due to the differences in the make of the devices and the protocols followed by different device vendors, the structure of the security layers of the connecting networks and the involvement of many low computational power devices, there are plenty of loopholes in IoT networks that hackers can exploit. Another point of concern is the integration of old devices that are either still fully functional or too expensive to replace, and which were not designed keeping network-based device integration in mind.

Among the most important aspects of IoT is the network architecture. Currently, TCP/IP is the most commonly used communication protocol for the Internet. But its implementation assumes some basic requirements—enough memory and computational power to generate and process TCP/IP data packets, and data format commonality among the connected devices. For IoT, both of these requirements may not be satisfied by the majority of the devices on the network. Different vendors have their own propriety versions of how the device reads and writes data. Also, devices like low-resolution cameras may not have enough embedded computing power and memory to handle complex calculations. Interfacing among such devices is an important aspect for successful IoT implementation. Many international forums have come up with device agnostic communication protocol frameworks that aim to resolve these issues. Once connected to the smart grid, for older devices with negligible computing power, data is simply forwarded to a central decision-making node, which crunches the data and issues appropriate instructions. The data is often also forwarded to the cloud, for the purpose of safe-keeping, and for running more extensive analysis tasks that can uncover long-term patterns of behavior that can be further exploited to improve operations and make them more intelligent. For newer devices, the development of concepts like federated learning and decentralized artificial intelligence means that even the smallest of devices with a given amount of memory and computational power is a smart powerhouse.

The importance of IoT arises from the fact that electro-mechanical devices are steadily getting smaller in size, with the improved computational capability and generating massive amounts of data every second. As time passes, there will only be more and more active, smart devices present in the environment around us. The extensive data gold mine generated by these connected, geo-strategically placed, stationary or moving devices, coupled with advanced data analysis and artificial intelligence algorithms, can be utilized to improve real-time efficiency and provide intuitive services, on the go. Collecting, storing and analyzing this data is today easier than ever before, due to the synergistic impact of the advances in networking, cloud storage and embedded computing domains. Meanwhile, innovative IoT-based applications are being developed in a collaborative manner and field-tested with an intention to be commercially launched in global markets.

If one keyword can be used to describe potential IoT applications, then it would be 'seamless.' The ultimate vision of the IoT community is to build an all-encompassing, integrated network of devices spread all over the globe—with multiple layers of authorization and privacy ensuring that the right information is disclosed to the right entities. As of now though, IoT is being applied in domain-specific silos, with semi-closed networks, both geographically and functionally. A 2016 survey by VMW revealed that only 45% of the businesses considered IoT highly relevant to their operations, but 72% thought that IoT will become very important for them in the next two to three years.

Let us take the example of the automobile industry. It has been among the earliest adopters of IoT—with the introduction of the concept of connected cars. This refers to a network of cars connected to other vehicles on the road, to the car manufacturers and service centers and to the user's many personal gadgets like the mobile phone. Connection to the manufacturer enables the cars makers to monitor the real-time status of the components in the cars and raise a concern to the car owner if the data analyzed says that maintenance is needed for a particular component. Predictive and preventive maintenance saves the user time as the maintenance schedule is tailored to the actual status of wear and tear of the car components. It allows the manufacturers to send the required replacement components to their service centers—enabling a more efficient, data-driven version of just-in-time inventory management. This also means that unanticipated car breakdowns are minimized. On the road, cars that communicate with each other ensure that their paths do not overlap—thereby reducing the risk of accidents. An intelligent, self-driving car can even get itself to the service center and gas station all by itself, without the owner's involvement.

Within the car manufacturing facility itself, IoT connected machines and robots can aid in waste reduction, improvement in precision engineering, and improvement in the overall safety of the factory by automating critical operations that just need to be monitored and controlled by the human operators from a safe room with the help of a computer screen. Thus, the potential applications of IoT in the automobile industry are immense, limited only by the power of imagination.

Another example is the oil and gas industry. With the help of robotic arms equipped with cameras and terrain sensors, exploration activities can be made less destructive, less expensive and faster. The usual process involves clearing the vegetation in the area suspected to have gas and oil reserves, drill massive holes at different locations and use cameras to explore these holes. If a drill breaks mid-operation, the drill arm needed to be extracted back to the surface and serviced. With the help of IoT, the exploration activities can be made more pinpointed. The terrain sensors on the robotic arm can judge in real time which kind of drill head to use and the status of wear and tear of the drill head. It can replace the drill head then and there itself, according to the requirements. The image feed from the robotic arm can be utilized to build a geological map of the underground terrain. Data from other sensors on the arm can be crunched to give the probability of finding gas in a particular direction, thereby streamlining drilling operations.

1.3 Artificial Intelligence (AI)

Over the last few years, artificial intelligence (AI) has regained its momentum due to recent advances in computing power, abundance in data availability and improvement in technical understanding. This has also led to the technology-driven organizations readily adopting advanced AI tools and techniques to solve challenging problems in managing their businesses. Artificial intelligence, as a concept, can be defined as smart systems with an ability to learn, think and solve problems using a set of computer algorithms and techniques ranging from speech recognition, pattern identification, deep learning capabilities, etc. With the advancements in computer science, machines are now able to perform cognitive functions which can mimic human actions such as learning, decision making and responding to events around them. Computerized machines of different sizes and shapes are now able to understand human speech signals, recognize optical characters and even compete with the best of human minds in the games of strategy such as chess and GO.

As traditional domains of AI research graduate from reasoning, knowledge representation and storage to advanced approaches involving cognition, learning and decision making, one of the classic challenges still happens to be revolving around how to make machines to understand ‘general intelligence’ or ‘common sense’ that most humans are generally, ‘gifted’ with. Drawing upon a seamless combination of various analytical scholarships such as mathematics, statistics and computational science with intuitive scholarships such as linguistics, psychology and philosophy, AI is helping machines to understand natural language processing, an ability to get a better sense of the surrounding situations and take decisions based upon the events happening to/around them. With natural language processing capabilities, IBM's Watson can interpret human dialogue and correctly understand their implied meanings. With machine learning and data analytical capabilities, Watson is able to learn from its past experiences and develop increasingly better solutions in areas where humans may sometimes make a mistake. For example, Watson is now being actively used to analyze MRI images to identify very small 'hemorrhages', which may not be easily visible to human eye. Such capabilities can be of tremendous help to the medical fraternity in areas of early-stage detection of potentially life-threatening diseases.

The literature in AI discusses various formal definitions of the term 'Artificial Intelligence.' The definitions vary along two dimensions, thought process(Think) and behavior(Act) of machines being developed, and the success of the machines being built are measured on two aspects, either human-like or being rational. So, historically there have been definitions on four approaches.

1. Machines that think like humans
2. Machines that act like humans
3. Machines that think rationally
4. Machines that act rationally.

The early history of AI says, although the dream of creating intelligent systems or machines existed for many centuries, the field of artificial intelligence has taken birth at a conference held at Dartmouth College in 1956. Marvin Minsky and John McCarthy organized the conference, and McCarthy coined the term 'Artificial Intelligence' for the proposal to obtain funding for the conference. Herbert Simon and Allen Newell were among the attendees of the conference, and these four people together were considered Fathers of AI. These four people started AI laboratories at premier research universities like MIT, Stanford and Carnegie-Mellon. Though the name artificial intelligence had few controversies, even among the researchers in the area, it eventually was accepted by the community.

Artificial Intelligence is inter-disciplinary field that heavily depends on different domains like computer science, cognitive science, engineering, ethics, linguistics, logic, mathematics, natural sciences, philosophy, physiology, psychology and statistics.

As we have seen earlier, there are many subtopics in the field of AI, but these topics generally come under one of the following technical research areas of AI. Each of the research areas below involves solving problems that are as holistic as solving the entire problem of AI in building an intelligent computer:

1. Natural language
2. Problem solving and search
3. Knowledge representation and reasoning
4. Learning
5. Vision
6. Robotics

Let us discuss few details and applications in each of these areas.

1. Natural Language:
   This area is basically the overlap of AI with computational linguistics, which basically aims at understanding how humans use their native languages and to be able to make a program that can use our languages at the same level of competence as a native speaker. Natural language understanding is one of the hot research areas today, as a complete natural language comprehensible machine can speak to humans, understand the nuances of the language used, context and emotion behind the text to strike a conversation seamlessly. So, the subtopics in this area include argument comprehension, natural language generation, semantic understanding of the language, machine translation, question answering, speech recognition, story analysis, lexical semantics, systemic grammar, etc.
2. Problem Solving and Search:
   This is a general area of AI concerned with discovering the solution to a problem. The main characteristic of this area lies in the approach to be taken in solving a problem, to find the best possible solution. This is of critical importance, as any problem can be solved in multiple approaches. The machine being built should be made intelligent enough, to choose the optimal and best path while solving the problem at hand. So, the area of problem solving is very much connected to search, as there is a need for computer to filter the best solution from the available solutions in search space. The subtopics here include distributed problem solving, heuristics, search algorithms, depth first, best first, branch and bound, simulated annealing, etc.
3. Knowledge Representation and Reasoning:
   Knowledge representation is concerned with building formal symbolic languages used to represent the knowledge (data) used by machines and the data structures used to implement these formal languages. The stored knowledge is intended to produce additional explicit knowledge, and this is called reasoning. These two things, knowledge representation and reasoning, are together necessary and sufficient to give the machine general intelligence. The topics under knowledge representation include belief network systems, dynamic memory, logic, semantic networks, etc., whereas topics under reasoning include Bayesian inference methods, causal reasoning, common sense, spatial reasoning, temporal reasoning, graph-based methods, truth maintenance, etc.
4. Learning:
   This is usually believed to be the critical characteristic of intelligent machine, and this is always the easy way to produce intelligent systems. This branch of AI is usually termed as machine learning, which is dealt separately, as a topic in its own, as it is of huge relevance and importance these days. The topics here include machine learning paradigms, supervised learning, semi-supervised learning, unsupervised learning, reinforcement learning, deep learning, learning theories, etc.
5. Vision:
   This research area deals with understanding how human vision works and trying to replicate the same in building intelligent systems that can understand the images and interpret them. This is emerging as a separate field, in the name

of computer vision. The research topics here include optical character recognition, intelligent character recognition, color vision, object detection, image understanding, stereo vision, captioning images, etc.

6. Robotics:
   This area is concerned with building machines that can move in the physical world and that can manipulate other objects in the world. The topics here include motion analysis, robot control systems, sensors fusion, range data analysis, robot manipulators, etc.

1.4 Robotics

Taking a peek into the history, the genesis of robots was observed during the Industrial Revolution in the late eighteenth century which witnessed the advent of mechanical looms then mechanical arm for spray painting, crane with motorized grippers and humanoid in the recent years. Primarily industry felt the need for the self-actuated machines to avoid humans from performing dangerous jobs or where it was practically impossible to physically reach work areas and also in those jobs where high amount of precision was required. In fact, on the other hand, these machines were helpful in jobs which were repetitive, boring, stressful or labor-intensive or menial tasks that humans wished to avoid. Since these machines behaved like a pseudo-worker, they were named robots, which meant servant or worker in Czech. These machines soon started finding wider industrial applications, and the need for appropriate conceptualization and regularization in usage was felt. Robot Institute of America carried out extensive work in the field of robotics. In 1979, they defined robot to be designed to perform a variety of task which is controlled via software program and is reprogrammable and multifunctional. However, in the present scenario, the understanding of the capabilities of robots has been altered whereby they can manipulate objects in the physical world, sense the information about the physical world and make real-time decisions based on the information or ask for more information. They are also expected to demonstrate friendly interface with humans and should require very less or almost no human supervision.

Robots aid quality improvement, better cost-effectiveness, give the industry the flexibility to change, improve the working environment and better productivity due to less material wastage, labor turnover and accidents. Extensive usage of robots across various industries has triggered a body of knowledge called robotics which deals with the engineering science and technology of robots, and their design, manufacturing, application and structural disposition. Though the robots give a wide range of benefits, it also poses few challenges such as reduction of human labor leading to loss of livelihood, expensive setup costs and lack of the human rationale which affects decision making in critical situations.

The technological advancements in the current scenario have opened the gates for wider applications of robots irrespective of the domains. Few of them are listed below:

1. **Space Robotics**: Development of general-purpose machines, which are capable of surviving extreme conditions and rigors of the outer space environment. They are designed to perform multiple tasks, including unanticipated tasks, within a broad sphere of competence such as planetary exploration. They can be partially controlled by humans from local control console or by astronauts in the shuttle. Space robots can perform tasks with higher cost-efficiency and also on an accelerated schedule. Using space robots lowers the risk with increased performance when compared to humans.
2. **Robotics in Hazardous Applications**: Hazardous environments can be typically characterized by radiological or toxicity dangers to potential explosions. Remotely operated systems are used to allow the human to operate in safe environments thereby reducing the risks caused due to the hazardous environment. These systems have been routinely used in explosives disposal, security operations and handling dangerous biological materials. They are also used for manufacturing chemicals and explosives.
3. **Underwater Robotics**: This technology is futuristic to aid ocean exploration and utilization. Subsea oil and gas factories, exposed aquaculture and deep-sea mining are the potential applications. Underwater robotics is expected to increase the level of autonomy and reduce the dependency on surface support for cost-efficient underwater operations. However, there are many challenges such as robustness against extreme weather condition, energy supply and communication disruptions which need to be championed before these futuristic expectations are realized.
4. **Virtual Exertions**: The virtual reality technology developed so far concentrates on giving a real-life experience to the user. However, the user cannot control the happenings in the scene and can exercise his control within a limited frame. The user experience can be improved through virtual exertions, which aims at evoking the sense of exerting forces in virtual reality using gestures and muscle activity. This can also be used to research and train individuals to perform strenuous tasks under potentially safer conditions.
5. **Robotics in Security and Military Applications**: Moving from detecting and exploding inground mines to ground-based surveillance using robotics, military application of robotics is exploring new horizons by initiating ground-based attacks, although at a debatable stage at the moment. These machines are equipped with machine guns and camera, thermal imaging and laser range finders capable of detecting intruders. Robot-based warfare may soon change the skills which the military seeks in future soldiers.

The above-discussed applications are just a glimpse of how robotics can greatly impact our lives in all possible ways. Since the advent of first industrial robot by General Motors, technology intensity and its subsequent applications have undergone a sea change. Although robotics allows the integration of ideas from various disciplines in engineering such as electrical engineering, mechanical engineering and computer science to design, develop and deploy robots which can perform multiple tasks as per the command, it constantly strives to aid, enhance and in certain cases even replace human actions. Such capabilities warrant the use of robots in situations which are deemed ‘dangerous’ for humans such as bomb detection and deactivation, risky manufacturing areas and construction sites. Efforts are increasingly moving toward the direction of replicating human cognition and behaviors such as speech, movement and even appearance. Research efforts are also being undertaken to keenly observe and understand ‘nature’ and the various living and non-living organisms around us in order to contribute toward nature-inspired robotics. This might help in creating robots which do not damage the nature while executing their commands. It has been speculated that robots will soon be able to assist, if not outperform humans in real-life decision making in complex situations where human values, morals and ethics are at stake.

1.5 Cloud Computing

Cloud computing facilitates shared access to large volume of computing resources and applied services like data storage, etc., over the Internet and in doing so eliminates the need for physical ownership of those resources at an individual user level. This allows organizations to benefit from economies of scale and thus reduce their computing infrastructure and maintenance costs. It does require an interrupted, stable and ubiquitous access to high-speed Internet. Service providers operating in this domain are trying to influence their clients into restructuring their business model and shift to cloud-based operating model in order to minimize their expenditure on computing and networking infrastructure. Clients may view this as ‘Software-as-a-service’ model and make use of it as per their requirements without physically owning the computing and storage resources. Cloud service providers generally operate on a ‘pay-as-per-use’ model, and clients should be well aware of the pricing models.

Cloud computing is a way of handling information technology infrastructure in a ubiquitous way. It helps in accessing large volume of computing resources over the Internet on a shared basis and results in the need for physical ownership of the resources at a user level. It allows organizations to benefit from economies of scale and thus reduce their computing infrastructure and maintenance costs.

It is very important in today’s connected world. Some of the reasons are mentioned below:

1. Cost
   Cloud computing reduces the capital expenditure drastically as it eliminates the need of buying hardware and software at each site.
2. Speed
   Another aspect of cloud computing is the speed as even vast amount of data can be processed in minutes providing businesses an edge in terms of speed of delivery.
3. Global scale

One of the best advantages of using cloud computing is that it can be scaled elastically and it provides flexibility in terms of computing power, storage and bandwidth.

4. Productivity
   Cloud computing helps in removing the need for many time-consuming tasks such as hardware setup and software patching thus saving time for the IT teams for more important business goals.
5. Performance
   Having cloud computing services on a worldwide network of data centers enables the organization to leverage economies of scale.
6. Reliability
   Data backup and disaster recovery is less expensive as data is mirrored at multiple redundant sites on the cloud provider's network making business continuity easier.

There are three types of cloud computing:

1. Infrastructure as a Service (IaaS)
   It is managed by a third party wherein infrastructures, such as hardware, software, servers and storage, are managed by the third-party vendor along with providing backup, security and maintenance.
2. Software as a Service (SaaS)
   With the help of cloud, we can use software online such as Internet browser. One can use the services of the software without even having it installed on their own system.
3. Platform as a Service (PaaS)
   With the help of cloud computing, users can develop, run and manage applications without having to get caught up in code, storage, infrastructure and so on.

1.5.1 Implementing Cloud Computing

Implementing cloud computing services involves four basic outsourcing phases:

1. Define strategy
   It includes the scope of IT, application and service, based on a transformation road map and aligned with business and IT. It also involves checking both functional and non-functional requirements.
2. Make a selection
   It involves researching on various options available and narrowing down to the one that fits the requirements. It also involves checking on the exit options available.
3. Deploy
   It implies integrating the service into the business, defining contingency exit strategy steps during implementation, governance, roles and responsibilities.

4. Manage and control
   It involves setting measurement parameters and key performance indicators (KPIs) for usage, performance and finance.

1.6 Machine Learning

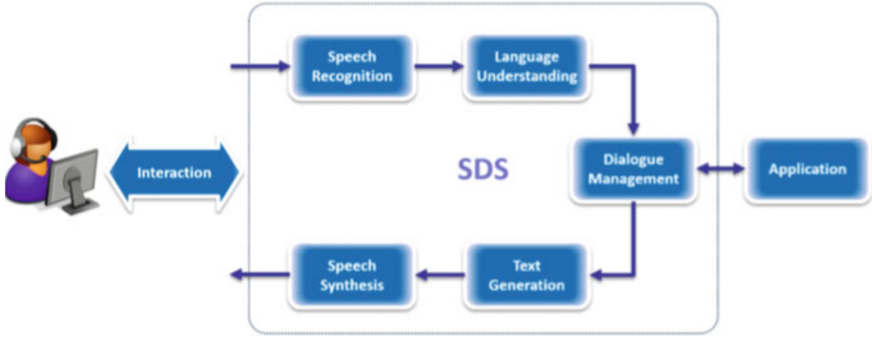
Machine learning algorithms leverage the power of statistical tools and techniques to empower computing systems with an ability to 'learn' about a particular task along with the associated data related to the task. They make use of computational learning theory, predictive and pattern recognition algorithms to facilitate data-driven decision making. The use of mathematical optimization techniques, machine learning can now be employed to generate new theories, processes and applications to improve human decision making. It employs complex mathematical, statistical and analytical modeling techniques to develop algorithms that can be used for prediction and behavior analysis. Such algorithms help decision makers in producing repeatable, reliable and robust results as well as revealing 'hidden insights' through cumulative understanding developed from historical data, proven models and analyzed trends in the past. Having said that machine learning is still in its initial stages and still has a long way to go in terms of delivering fail-proof results.

Deep Learning, under the umbrella of machine learning, is being heavily used in the industry to build intelligent systems that can assist humans in various tasks. The world is increasingly interacting with machines by talking to them, for example, when using Amazon's Alexa, Apple's Siri or Microsoft's Cortana. Automatic machine translation has been around for a long time; it is only now that deep learning methods are achieving top results, especially in automatic translation of text (language translation) and automatic translation of images. Large recurrent neural networks are being trained to learn the relationship between input sequences (audio, image, text) and labeled output sequences (transcription, translation, handwriting, etc.). Customized deep convolutional neural networks are being graphed and trained to achieve never seen results in object detection within the domain of image recognition. Layers of long short-term memory (LSTM) units on the top can make use of the learnings for detecting objects to generate coherent sentence descriptions (Automatic image caption generation). The range of applications, where top results are being achieved, extend to the fields of data-driven predictive advertising, predictive finance, medical imaging and biomolecular computations, predicting earthquakes, etc.

One of the problem statements of interest is the spoken dialogue system, which comprises mainly five modules:

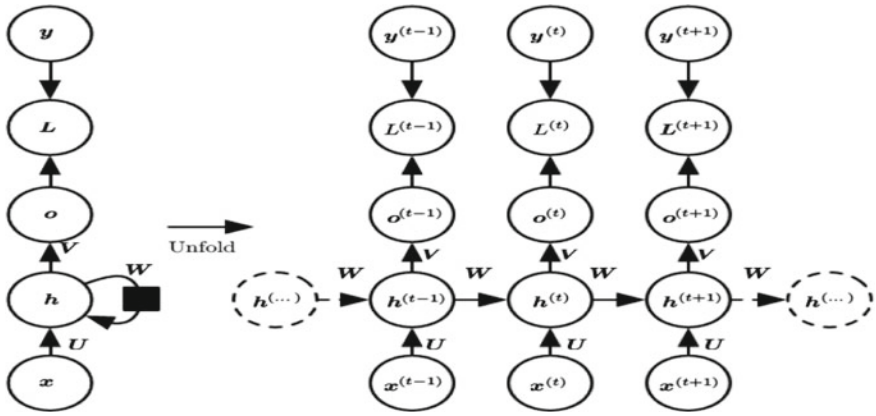
(a) Speech Recognition: Transcribes the speech to text;
(b) Language Understanding: Semantic interpretation, intent and entity recognition;
(c) Dialogue Management: Based on the intent and entity, the dialogue manager decides which system action needs to be executed;

(d) Text Generation: The system action is transformed into a sequence of words;
(e) Speech Synthesis: The sequence of words is spoken out loud by synthesizing audio.



Source User-centred Adaptive Spoken Dialogue Modelling, Dissertation, Stefan Ultes, ULM University

The first module, i.e., the automatic speech recognizer, is a classic example of a sequential input (audio) to sequential output (text) learning. The audio sequence's input representation is featured as a time series of energy levels in different frequency bands (Mel-frequency Cepstral Coefficients representation). The end-to-end speech to text conversion by deep recurrent neural networks (RNNs) outperforms the hybrid hidden Markovian neural net approach. The main advantage of using recurrent layers is that they are inherently deep in time since their hidden state at current time is a function of both, input at current time-step and all its previous hidden states as well. This helps to learn the useful context necessary for prediction on a sequence of audio input at each time-step. A simple RNN along with the output and loss layer looks like the following figure, when unfolded in time:



Source The Deep Learning Book, Ian Goodfellow, Yoshua Bengio, Aaron Courville

Another big advantage of using this discriminative, end-to-end neural-only approach to classification is that it removes the need for predefined alignment between input audio sequence of length T and output label sequence of length ($\leq T$), to create training targets for the network. The network outputs at each time-step are used to parameterize a differentiable distribution $\Pr(\pi|x)$ over all possible text sequences (π) of length T given the acoustic input sequence x. The next step is to calculate the log-probability log $\Pr(z|x)$ of the actual-text-target output sequence (z) of length $\leq T$, by using a dynamic programming algorithm, similar to the forward-backward algorithm in case of HMMs. These log probabilities can then be differentiated with respect to network weights using back propagation through time (BPTT), and hence, the whole network can be trained by optimizing with gradient descent. This technique of sequence–to-sequence classification where the alignment between the input and output is not pre-known comes under the framework of connectionist temporal classification.

Once the model weights and biases are learnt, a decoding algorithm (such as beam search decoding) is used along with an assisting language model, to do the transcription of a new audio example during inference. The language model helps constrain the search space of words in the unidirectional decoding process. It helps the speech recognizer figure out how likely a word sequence is, independent of the acoustics' ambiguity.

The performance measure that can be used on the validation and test set is the edit distance (ED(p, q)), i.e., the minimum insertions, substitutions and deletions required to convert sequence p to sequence q, for example, the Levenshtein distance.

Once the speech utterance is recognized correctly to text, detection of intent and semantic slot filling are the two major tasks within the language understanding framework for dialogue systems. The goal of intent detection is to assign an intent class from a predefined set of intent classes to a sequence of words. This can be applied by a variety of standard classifiers like support vector machines (SVMs), multiclass logistic regression or also by RNNs.

The goal of slot filling is to find a sequence of semantic labels $s = (s1, s2, \ldots, sM)$, one for each word $w = (w1, w2, \ldots, wM)$ in the utterance such that each word can be read as either filling one of the predefined required slots or carrying no semantic meaning, i.e.

$$s^* = \text{argmax}s\ (\Pr(s|w))$$

The response message generation requires to convert the abstract system actions (fetching of relevant data fillings from the database) to natural language. The generated text message can then be synthesized to speech, incorporating techniques to imitate human-like responses.

The evident challenges in the area of research in cognitive technologies are to find efficient ways to integrate all sub-modules while preserving the tractability and accuracy at each step. The humanoid, Sophia, a Saudi Arabian Citizen, is an embodiment of overcoming all such challenges.

1.7 Big Data Analytics

Big data analytics deals with the discovery, interpretation and the subsequent communication of insightful patterns in the seemingly complex, ambiguous and voluminous data. It heavily draws upon concepts from computer programming, statistical and operational research to assess the performance of a system producing large amounts of real-time data. Increasingly, technology-driven organizations are using it to understand, predict and overhaul their business performance. With the use of advanced computing devices and algorithms, big data algorithms have become robust enough to handle complex and voluminous data sets which traditional data sets cannot deal with. However, there are still certain challenges such as data collection, storage and analysis along with the efficiency and effectiveness of the searching, sharing and visualizing the data. All these major aspects have been briefly summarized in the five attributes of big data analytics: volume, velocity, variety, value and veracity of the data.

With such challenges ahead, experts in the field are extracting valuable information from the data and employing predictive analytical algorithms to generate game-changing insights for their clients. With the use of powerful data crunching algorithms, new set of relationships are being uncovered to combat lethal diseases, terror plots and criminal activities. Business-oriented organizations are harnessing the power of analytics to forecast their growth, emerging trends and future business opportunities. Retail organizations are using analytics software packages specially designed to manage store operations such as inventory management, stock optimization and consumer behavior. Marketing professionals are leveraging it for predicting sales, promotion modeling and improve their brand positioning. Financial experts are using it to track the markets, predict the stock prices and credit-risk analysis, etc. Scientists are using big data algorithms to solve complicated problems in astrophysics, biology and environmental research.

Big data analytics is not just the analysis of large and complex data; there are various aspects to it. There are several types of technologies that work together which helps in getting most value from the data. Some of these main technologies include: data management which covers the aspect of making the available data reliable and of the highest possible quality adhering to the quality standards throughout the organization; data mining is examining large amounts of data to discover patterns which can be used in further analysis, helping in answering complex business questions; Hadoop is a software framework used to run applications on clustered commodity hardware, its distributed computing model enables faster processing of the big data; in-memory analytics caters to decreasing the data prep and analytical processing latencies by the analysis of the data directly from the system memory; predictive analytics uses historical data to predict the likelihoods of the future outcomes by the use of statistical algorithms and machine learning techniques; and lastly, text mining deals with the analysis of the text data which is sourced through the emails, blogs, surveys, twitter feeds or any other social media data, and this data is then processed and analyzed through the machine learning and the natural language processing techniques.

There are various data mining techniques which help in the meaningful and application-based analysis of the big data. Association, classification, clustering, prediction and decision tree are some sort after data mining techniques. These techniques are being widely used in various domains such as travel and hospitality, retail, health care and public sector. Big data analytics gives the businesses a potential to collect customer data, analyze it and provide insights as to where the business opportunity lies, which segments to targets, what products to sell; similarly, in the healthcare industry, it helps in analyzing large amounts of patient records which help in providing immediate lifesaving diagnosis and treatment options. In the public sector, it helps in various aspects of governance, one of its examples being crime monitoring and prevention by effective monitoring and analysis of the crime-related data coming from previous crime scenarios, surveillance-related data, etc.

1.8 Conclusion

In this chapter, we briefly introduced and explained some of the 'next generation' smart technologies along with their potential areas of application. In the remaining part of the book, we attempt to present detailed insights about the manner in which some of these technologies are expected to impact key areas like security, agriculture, manufacturing, healthcare, transportation, education, retail, insurance, banking and financial services etc. in the Indian context. The book also throws light on the challenges posed while adapting these technologies accompanied by some recommendations to overcome them. Some important concerns related to the development and deployment of such large-scale technology platforms have also been discussed in a very simple and lucid language.