1 Cybernetic View of Robot Cognition and Perception

1.1 Introduction to the Model of Cognition

The word ‘cognition’ generally refers to the faculty of mental activities of human beings dealing with abstraction of information from the real world, their representation and storage in memory, as well as automatic recall [Patnaik et al., 2003a]. It includes construction of higher level percepts from low level information or knowledge, which is referred to as perception. It also does the construction of mental imagery from real instances for subsequent usage in recognizing patterns or understanding complex scenes. It includes various behaviors like sensing, reasoning, attention, recognition, learning, planning and task coordination, as well as the control of activities of human beings. Cognitive science is a contemporary field of study that tries to answer questions about the nature of knowledge, its components, development and uses [Matlin, 1984]. Cognitive scientists have the opinion that human thinking involves the manipulation of internal representation of the external world known as the *cognitive model*. Different researchers have investigated various models of human cognition during the last thirty years. Those were basically analytical and experimental psychology, and gradually scientists have tried to implement this knowledge in developing intelligent robots.

A very first model of a robot was developed by dividing the entire task into a few subtasks, namely *sensing, planning, task coordination* and *action,* which was popularly known as the *principle of functional decomposition*. These subtasks were realized on separate modules that together form a chain of information flow from the environment to the actuators, through sensing, planning and task coordination. The primitive model is poor at accommodating major components like perception, map building and world modeling. Subsequently, these functional modules were included in the robot model, by various researchers.

In 1986, Rodney A. Brooks was the first man to use the findings of ethological research, and to design a mobile robot. He published a seminal paper on the *subsumption architecture,* which was fundamentally a different approach in the development of mobile robots [Arbib, 1981]. He developed the *subsumption language* that would allow him to model something analogous to animal behaviors in tight *sense-act* loops using asynchronous finite-state machines. The first type of behavior for a robot was used to avoid obstacles that are too close and moving a little away or else standing still. Secondly, higher level behavior might be to move the robot in a given direction. This behavior would dominate the obstacle-avoidance behavior by suppressing its output to the actuators unless an object comes too close. The higher levels subsumed the lower levels, and therefore the name of the architecture was *subsumption architecture*. They were able to develop a robot, using simple sonar or infrared sensors that could wander around a laboratory for hours without colliding into objects or moving people. After this development, Brooks and his colleagues developed highly mobile robots, i.e. mobots, both wheeled and legged, which could chase moving objects or people and run or hide from light. Further, they can negotiate a cluttered landscape which might be found in a rugged outdoor environment.

During the 1990s, there were many developments such as HERBERT: a soda-can-collecting robot [Connell, 1990]; GENGHIS: a robot that learned to walk [Maes & Brooks, 1990; Brooks, 1989]; TOTO: a hallway-navigating robot [Matric, 1992]; and POLLY: a tour-guide robot [Horswill, 1993]. The idea was to build up capability in the robot through behaviors that run in parallel to achieving possible alternative goals. The behavior in these robots was able to execute various actions on a priority basis or to achieve various goals within the cycle time. After this development, the world model was distributed among the types of behaviors with only relevant parts of the model being processed for each behavior. The generation of simple plans for path planning and the compilation of the result of actions could be done before run time by using these types of behaviors.

Today, researchers are trying to develop intelligent machines after a careful and meticulous review of human cognition, soft computing tools and techniques. But there are still open problems in these areas of machine learning and perception, which are being investigated using many alternative approaches. This research work mainly aims at studying various techniques of perception and learning, using the cognitive model, and their applications in mobile robots. Detailed programs have been provided in the respective chapters.

1.1.1 Various States of Cognition

Let us introduce a model of cognition that includes seven mental states, namely *sensing and acquisition, reasoning, attention, recognition, learning, planning, action and coordination* and their transitions along with *cognitive memory,* i.e. *LTM* and *STM*, as shown in Fig. 1.1. There are three cycles embedded in the model, namely the *acquisition cycle, the perception cycle* and *the learning and coordination cycle*, which describe the concurrent transition of various states.

The acquisition cycle consists of two states namely *sensing and attention* along with *short term memory* (STM) and *long term memory* (LTM), whereas the *perception cycle* consists of three states namely *reasoning, attention and recognition* along with LTM. The *learning and coordination* cycle consists of LTM along with three states namely, *learning, planning* and *action*. The various states are explained as follows.

Sensing and acquisition: Sensing in engineering science refers to reception and transformation of signals into a measurable form, which has a wider perspective in cognitive science. It includes preprocessing and

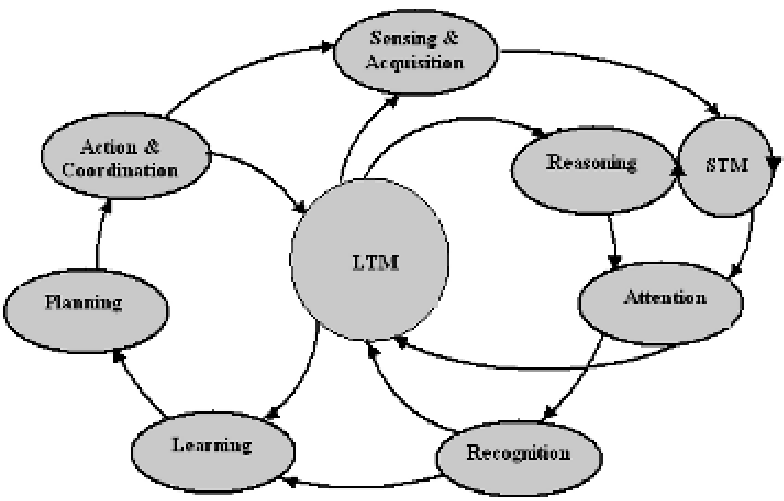

Fig. 1.1. Three cycles namely acquisition, perception and learning and coordination, with their states in the model of cognition. LTM = Long Term Memory; STM = Short Term Memory

extraction of features from the sensed data, along with stored knowledge in LTM. For example, visual information on reception is filtered from undesirable noise, and the elementary features like size, shape, color are extracted and stored in STM [Borenstain, 1996].

Reasoning: Generally this state constructs high level knowledge from acquired information of relatively lower level and organizes it in structural form for efficient access [Bharick, 1984]. The process of reasoning analyses the semantic or meaningful behavior of the low level knowledge and their association [Chang, 1986]. It can be modeled by a number of techniques such as commonsense reasoning, causal reasoning, non-monotonic reasoning, default reasoning, fuzzy reasoning, spatial and temporal reasoning, and meta-level reasoning [Popovic et al., 1994].

Attention: This is responsible for the processing of a certain part of the information more extensively, while the remaining part is neglected or suppressed. Generally, it is task-specific visual processing which is adopted by animal visual systems [Matlin, 1984]. For instance, finding out the area of interest in a scene autonomously is an act of attention.

Recognition: This involves identifying a complex arrangement of sensory stimuli such as a letter of the alphabet or a human face from a complex scene [Murphy et al., 1998]. For example, when a person recognizes a pattern or an object from a large scene, his sensory-organs process, transform and organize the raw data received by the sensory receptors. Then the process compares the acquired data from STM with the information stored earlier in LTM through appropriate reasoning for recognition of the sensed pattern.

Learning: Generally speaking, learning is a process that takes the sensory stimuli from the outside world in the form of examples and classifies these things without providing any explicit rules [Winston, 1975]. For instance, a child cannot distinguish between a cat and a dog. But as he grows, he can do so, based on numerous examples of each animal given to him. Learning involves a teacher, who helps to classify things by correcting the mistake of the learner each time. In machine learning, a program takes the place of a teacher, discovering the mistakes of the learner. Numerous methods and techniques of learning have been developed and classified as supervised, unsupervised and reinforcement learning [Baldi, 1995; Carpenter et al., 1987; Lee et al., 1997].

Planning: The state of planning engages itself to determine the steps of action involved in deriving the required goal state from known initial states of the problem. The main task is to identify the appropriate piece of knowledge derived from LTM at a given instance of time [McDermott et al., 1984]. Then planning executes this task through matching the problem states with its perceptual model.

Action and coordination: This state determines the control commands for various actuators to execute the schedule of the action plan of a given problem, which is carried out through a process of supervised learning [Maes & Brooks, 1990]. The state also coordinates between various desired actions and the input stimuli.

Cognitive memory: Sensory information is stored in the human brain at closely linked neuron cells. Information in some cells may be preserved only for a short duration, which is referred to as short term memory (STM). Further, there are cells in the human brain that can hold information for quite a long time, which is called long term memory (LTM). STM and LTM could also be of two basic varieties, namely *iconic memory* and *echoic memory*. Iconic memory can store visual information whereas the echoic memory deals with audio information. These two types of memories together are generally called *sensory memory*. Tulving alternatively classified human memory into three classes, namely episodic, semantic and procedural memory [Tulving, 1987]. Episodic memory saves the facts as they happen; semantic memory constructs knowledge in structural form, whereas procedural memory helps in taking decisions for actions.

1.1.2 Cycles of Cognition

Acquisition cycle: The task of the acquisition cycle is to store the information temporarily in STM after sensing the information through various sensory organs. Then it compares the response of the STM with already acquired and permanently stored information in LTM. The process of representation of the information for storage and retrieval from LTM is a critical job, which is known as *knowledge representation.* It is not yet known how human beings store, retrieve and use the information from LTM.

Perception cycle: This is a cycle or a process that uses the previously stored knowledge in LTM to gather and interpret the stimuli registered by the sensory organs through the acquisition cycle [Gardener, 1985]. Three relevant states of perception are *reasoning, attention* and *recognition,* and are generally carried out by a process of unsupervised learning. Here, we can say that the learning is unsupervised, since such refinement of knowledge is an autonomous process and requires no trainer for its adaptation. Therefore, this cycle does not have “Learning” as an exclusive state. It is used mainly for feature extraction, image matching and robot world modeling. We will discuss human perception in detail in the next section, with applications.

Learning and coordination cycle: Once the environment is perceived and stored in LTM in a suitable format (data structure), the autonomous system utilizes various states namely *learning, planning* and *action and coordination* [Caelli & Bischob, 1997]. These three states taken together are called the *Learning and Coordination Cycle*, which is utilized by the robot to plan its action or movement in the environment.

Cognition, being an interdisciplinary area, has drawn the attention of researchers of diverse interest. Psychologists study the behavioral aspects of cognition and they have constructed a conceptual model that resembles the behavior of cognition with respect to biological phenomena. On the other hand, the engineering community makes an attempt to realize such behavior on an *intelligent agent* by employing AI and soft computing tools. The robot as an intelligent agent receives sensory signals from its environment and acts on it through its actuators as well as sensors to execute physical tasks.

This book covers techniques for feature extraction, image matching, machine learning and navigation using the cognitive method. A mobile robot senses the world around it through different transducers, such as ultrasonic sensors, laser range-finders, drives and encoders, tactile sensors, and mono or stereo cameras. The sensory information obtained by a robot is generally contaminated with various forms of noise. For instance, ultrasonic sensors and laser range-finders sometimes generate false signals, and as a consequence determination of the direction of an obstacle becomes difficult. The acquisition cycle filters the contaminated noise and transfers the noise-free information to the LTM. The perception cycle constructs new knowledge of the robot’s environment from the noise-free sensory information. The learning and coordination cycle executes various tasks assigned to it. These three cycles, along with their states, are utilized in modeling various techniques. The proposed work borrows the ideas from the model of cognition and implements them through various soft computing tools. The subsequent sections discuss issues like machine learning and perception in detail.

1.2 Visual Perception

Vision is the most powerful sense organ of human beings and it is also the key sensory device for a mobile robot. So far, not much progress has yet been achieved in visual perception of a mobile robot due to limitations in hardware and software. Visual processing requires specialized hardware and cameras, which are quite large to fit into the mobile robot. Secondly, traditional software for vision processing is very poor in quality, because it requires complete analysis of the entire scene even to recognize a minute object. Further, detection of an obstacle in front of the robot using stereovision takes a longer time, which is not at all permissible for online navigation.

As there is a shift in paradigm towards the behavioral model, researchers have started examining animal models for both motor control and perception. The research findings reveal that the frog uses simple visual motion detection to catch flying prey and bees depend on specific aspects of the color spectrum for their search. Psychological study indicates that the human visual system supports very simple behavior. Low resolution peripheral vision is used to watch for indications of motion, for instance collision with looming objects, whereas the high resolution fovea is used to gather information for reasoning about an object. Human vision does not perceive everything in all its color, motion and temporal dimensions at one time but direct attention is given to a very narrow portion of the visual field based on the task they are performing. As a result of this study, the paradigm in vision is shifted to a philosophy where perception exists to support the behavior of robots [Murphy et al., 1998].

Dickmann studied two major principles of perception during the 1980s. The first one was about the evolving process and internal representation of the world, which is known as Schopenhauer's idea of perception, and the second one was Kant's theory of the *true reality of perception*. With these two principles, Dickmann could represent various systems, including real time constraint using the notion of space and time [Zavidovique, 2002]. Human beings acquire knowledge of their surroundings unconsciously during their first years of crawling, then walking and reacting. Let us discuss a little more, the physiology and anatomy of the human visual system, which may help in understanding robot vision.

1.2.1 Human Visual System

The human visual system converts energy in the visible spectrum into action potentials in the optic nerves. The wavelength of visible light ranges

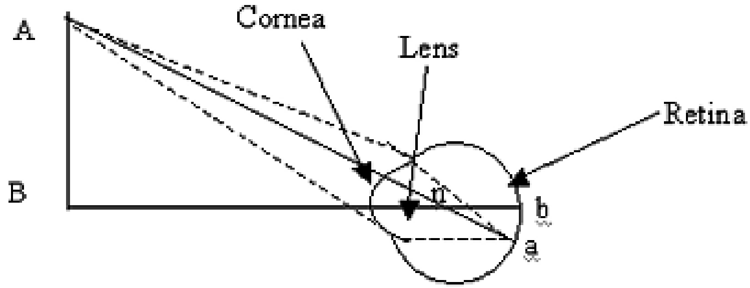

Fig. 1.2. A schematic diagram of human vision. n = nodal point, AnB and anb are similar triangles. In this diagram, the nodal point is 15 mm from the retina. All refraction is assumed to take place at the surface of the cornea, 5 mm from the nodal point. The dotted lines represent rays of light diverging from A and refracted at the cornea so that they are focused on the retina at 'a'

from approximately 390 nm to 720 nm. In the human eye, light is actually refracted at the anterior surface of the cornea and at the anterior and posterior surface of the lens. The process of refraction is shown schematically in Fig. 1.2.

The images of objects in the environment are focused on the retina. The light rays striking the retina generate potential changes that initiate action potentials on photosensitive compounds such as rods and cones. When the compound absorbs light, the structure changes and triggers a sequence of events that initiate neural activity. The human eye contains about 130 million rods and approximately 8 million cones. Rods are monochrome receptors of light, and cones are sensitive to color, which are different parts of the spectrum [Seculer et al., 1990]. The distribution of rods and cones in the retina is very irregular. The detailed spatial representation of the retina is in the form of electrical responses transmitted to the lateral geniculate nucleus (LGN) via retinal ganglion cells (RGC). Subsequently, LGN projects a similar point-to-point representation on the visual cortex, where these electrical responses produce the sensation of vision.

1.2.2 Vision for Mobile Robots

Vision is the fundamental part of perception in intelligent robots, the way it is for humans. The perception objective depends on three basic system qualities, namely *rapidity, compactness* and *robustness*. Active vision is the theoretical analysis of the vision process originated by Aloimonos et al. [Aloimonos, 1987; Fermuller & Aloimonos, 1993] to optimize the 3D reconstruction; and animate vision [Ballard, 1991], which is based on human perception analysis. Perception is an essential and most useful process for a mobile robot. To operate the mobile robot in unknown and unstructured environments, the robot must be able to perceive its environment sufficiently so as to operate safely in its environment. It is clear from study of the animal visual system that the animals concentrate on a specific part of the image received through their visual system, or in other words animal visual system does the task-specific visual processing. This philosophy has been borrowed to develop a task-oriented approach for sensing, planning and control.

In the realm of autonomous control, let us briefly mention visually guided control systems and the role of computer vision in autonomously guided robot systems. Hashimoto has introduced a closed-loop control system for visually guided manipulators [Hashimoto, 1999]. Visual information about tasks and the environment is essential for robots to execute flexible and autonomous tasks. A typical task of autonomous manipulation is to track a moving object with the robot hand, based on information from a vision sensor. To carry out this job, Hashimoto introduced a feedback loop with vision sensor, which can track a moving obstacle (Khepera robot) that moves on the floor. Sullivan et al. [Sullivan, 1999] introduced a system that tracks a moving, deformable object in the workspace of a robotic arm fitted with a camera. For their experiment, they used a figure-ground approach for object detection and identification. The figure-ground methodology allows pixels to be identified as objects or background pixels, a distinction which is useful during the initial placement of the active deformable model.

The evolution of machine perception has taken place during the last few decades [Zavidovique, 2002; Merlo et al., 1987]. The first system for signal analysis and pattern recognition was designed on a commercial computer in which a special-purpose interface was built for data acquisition using standard cameras and microphones. At that time, serious limitations were faced in signal transfer rate and core memory. Both sound and image preprocessing was developed during the 1970s using ad hoc hardware and suitable algorithms. Later, to improve the machine perception strategies, the link between sensing and processing was operated in a closed loop to obtain so-called *active perception*. *Active perception* is the study of perception strategies including sensor and signal processing cooperation, to achieve knowledge about the environment [Merlo et al., 1987]. Both high level and low level image processing need to consider perceptual information in order to reduce uncertainty.

Initially, the aim of the robot vision designer was to build the simplest possible system that was necessary to solve a given task and use its performance to improve its architecture. Horswill has developed a low cost vision system for navigation called the POLLY System, where the development team has used active, purposeful and task-based vision [Horswill, 1993], which computes the specific information needed to solve specific tasks. Murphy has introduced another model of sensing organization called action-oriented perception for a mobile robot, with multiple sensors performing locomotive tasks [Murphy, 1998], and work on robot vision and perception is still continuing as a major research topic.

1.3 Visual Recognition

As mentioned earlier, recognition is the important component of perception. Human beings are able to perceive and move around in a dynamic world without any difficulty but robot vision requires a large amount of computing resources and background knowledge for a very limited and static environment. There are many hypotheses of representation for various shape recognition techniques [Caelli & Bischob, 1997]. The representation contains information about the shape and other properties such as color, dimensions and temporal information, including a label, i.e. a name of the object. The objective is to retrieve the label correctly during the recognition process. Representations are stored in LTM as a set of separable symbolic objects or class of objects.

Recently developed models do not use a representation that is a direct replica of the retinal stimulation. Rather, they introduce the representation which deals with invariant properties of different objects in various positions, sizes, rotations, and even under different lighting conditions. During recognition, the captured image corresponding to an unknown object is converted to the same format and representations, which provides the best match using the same form of similarity measure. Each theory may have different assumptions regarding various parameters, such as:

- Type of representation, i.e. feature space, predicates, graph, etc.
- The number of representations per object, i.e. one 3D representation or multiple 2D representations from different viewing positions
- The number of classes for mapping into representations
- Inclusion of spatial relationships between objects and their component parts
- The amount and type of preprocessing given to the initial retinal image matching algorithm.

The primary issues in the visual recognition process are representations and search, which means how to develop an appropriate representation for the objects and then how to search them efficiently for a match at the time of recognition. Here are some representations used in traditional theories.

1.3.1 Template Matching

Template matching is the simplest form of representation in which a replica of the retinal stimulation pattern projected by a shape is stored in LTM. The recognition process compares all stored object templates with the input array by selecting the best match based on the ratio of matching to non-matching objects [Briscoe, 1997].

There are many problems with this method which prevent recognition, such as: (i) partial matches can give false results, for example comparing 'O' with 'Q'; (ii) any change in the distance, location or orientation of the input object in relation to the corresponding stored object will produce a different pattern; and (iii) any occlusion, shadow or other distortion of the input object may also produce inaccurate matching.

Some systems, for example, attempt to compensate for these problems by storing multiple templates, each recorded at various displacements, rotations and sizes. However, the combinatorics of the transformations usually prove to be cumbersome. The option of rotating, displacing or scaling of the input pattern to a canonical form before matching is also not feasible, as the required transformations cannot be known until the object is recognized. But the major limitation of this representation is that it is only appropriate for an object recorded in isolation. The template models are not useful at all if multiple objects are present in a scene, because the method is unable to determine which parts belong to which object.

1.3.2 Feature-Based Model

Instead of storing templates for entire shapes, the feature-based model utilizes a series of *feature detectors*. Generally the features included are of a geometric type such as vertical and horizontal lines, curves and angles. Feature detectors may be used either at every position in the input array, or may be used for the global image. In case of multiple feature detectors, the degree of matching is estimated for the target feature with respect to each section of the input array. The levels of activation for each feature may be summed up across the input array by providing a set of numbers for each feature. This list of numbers in the form of a vector of weights for different features is used as the stored representation of the object. The objective is to define the shape of the object with invariant features, which are independent of locations. The process of recognition consists of finding the best match between the stored representations and the levels of activation of the feature detectors in the input image.

1.3.3 Fourier Model

In the Fourier model, a two-dimensional input array is subjected to a spatial Fourier analysis. In this model, the original array is decomposed into a set of spatial frequency components of various orientations and frequencies in the form of sinusoidal waveforms. The amplitude and phase are both recorded for the spectrum of spatial frequencies and angles. Thus the original image is represented as the sum of the spatial frequency components and this transform retains all the details of the original image. The feature of this model is that it gives no restriction on angles, frequencies and no computational problems, such as *aliasing*. Even though the amplitude spectrum contains shape information and the phase spectrum contains position information, there is no method available for combining this information in order to locate a particular object at a particular location.

Each shape is stored in the memory in the form of its Fourier transform and its recognition is done by matching this with a similarly transformed input image. This model separates information about sharp edges and small details from other information pertaining to gross overall shape. Techniques such as edge detectors and convolutions may be used to extract these different details of the original image. The advantage of this model is that it can also match blurred edges, wiggly lines and other slightly distorted images.

1.3.4 Structural Model

The structural model contains information about the relative positions and relationships between parts of an object. This structural description is stored in memory in the form of a data structure such as a list or tree or graph of predicates. The representation is often depicted as a graph, where nodes correspond to the parts or the properties, and the edges correspond to the spatial relations [Minsky, 1975]. The advantage of structural representation is that it factors apart the information in a scene without losing any part of it. This model enables us to represent the object with the help of a list of labels and also their relative position and orientation with respect to the human observer. Various spatial reasoning operations may be performed by specifying the shape, location, orientation and spatial relationship of one set of objects with other objects in another set. The recognition process can be improved by including statistical and logical operations. The use of structural descriptions appears to be preferred because of computational convenience.

1.3.5 The Computational Theory of Marr

The work of David Marr [Marr, 1982] is one of the best examples of the computational approach to the recognition problem, which is the most influential contemporary model of 3D shape recognition. Marr introduced the need to determine edges of an object and constructed a 2½ D model, which carries more information than 2D but less than a 3D image. Thus an approximate guess about the 3D object can be framed from its 2½ D images.

1.4 Machine Learning

Since the invention of the computer there was always the question of how to make them learn. If we could understand how to program them to learn, i.e. to improve automatically with experience, it would have been a great achievement. A successful understanding of how to make them learn would open up many new uses of computers and new levels of competence and customization. Further, a detailed understanding of machine learning might lead to further investigation of human learning ability and disabilities. Machine learning algorithms have been investigated by a number of researchers. These are effective for certain types of learning tasks and as a result a theoretical understanding of learning started to emerge. Broadly speaking learning means any computer program that improves its performance for some tasks through experience.

1.4.1 Properties and Issues in Machine Learning

The formal definition of learning is: *A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T as measured by P improves with experience E* [Mitchell, 1997]. To give a specific example of learning, let us consider an example of an autonomous driven vehicle. The machine learning method has been used to train a computer-trained vehicle to steer correctly when driving on a variety of roads. For instance, the

ALVINN system [Pomerleau et al., 1989] has used its learned strategies to drive unassisted at 70 miles/hour for 90 minutes on public highways among other cars. Similar techniques have possible applications in many sensor-based control systems. This learning problem of autonomous driven systems can be formally defined as

Task T: driving on public highways using vision sensors.
Performance measure P: average distance traveled before an error occurs as judged by a human observer.
Training experience E: a sequence of images and steering commands recorded while observing a human driver.

There are various factors to be considered while designing a learning system, which are given as follows.

Choosing the training experience: The type of training experience available can have a significant impact on the success or failure of the learner. There are a few key attributes to contribute to the success of the learner. The first key is whether the training experience provides direct or indirect feedback regarding the choices made by the performance system. The second attribute of the training experience is the degree to which the learner controls the sequence of training examples. The next attribute is how well it represents the distribution of examples over which the final system performance P must be measured. Generally speaking, learning is most reliable when the training examples follow a distribution similar to that of future test examples.

Choosing the target function: This is to determine exactly what type of knowledge will be learned and how this will be used by the performance program. In other words, the task is to discover an operational description of the ideal target function or an approximation to the target function. Hence, the process of learning the target function is often called the function approximation.

Choosing the representation of a target function: A representation of the target function has to be described for the learning program to learn. In general the choice of representation involves a crucial trade-off. On one hand, a very expressive representation has to be picked up in order to obtain an approximate representation function as close as possible to the ideal target function. On the other hand, the program will require more training data for the more expressive representation. Therefore, it is advisable to choose an alternative representation, which can accommodate a broad state of training data and function.

Choosing a function approximation algorithm: In order to learn the target function, a set of befitting training examples is required. For instance, in the case of robot navigation a set of sensory readings and robot movement forms a set of training patterns for learning. After the derivation of the training examples from the training experience available to the learner, the weights have to be adjusted to best fit these training examples.

One of the useful perspectives on machine learning is that it involves searching every large space of possible hypotheses to determine one that best fits the observed data and any prior knowledge held by the learner. The machine learning algorithm has proven itself useful in a number of application areas, such as:

- poorly understood domains where humans might not have required knowledge to develop effective algorithms, for example human face recognition from images;
- domains where the program must dynamically adapt to changing conditions, for example controlling manufacturing process under changing supply stock;
- they are especially useful in data mining problems where large databases may contain valuable implicit regularities that can be discovered automatically.

1.4.2 Classification of Machine Learning

Machine learning can be broadly classified into three categories: (i) *supervised learning*, (ii) *unsupervised learning* and (iii) *reinforcement learning*. *Supervised learning* requires a trainer, who supplies the input–output training instances. The learning system adapts its parameters using some algorithms to generate the desired output patterns from a given input pattern. But, in the absence of trainers, the desired output for a given input instance is not known, and consequently the learner has to adapt its parameters autonomously. This type of learning is termed *unsupervised learning*. There is a third type of learning, known as *reinforcement learning*, which bridges the gap between supervised and unsupervised categories. In reinforcement learning, the learner does not explicitly know the input–output instances, but it receives some form of feedback from its environment. The feedback signals help the learner to decide whether its action on the environment is rewarding or punishable. The learner thus adapts its parameters based on the states of its actions, i.e. rewarding or punishable. Recently, a fourth category of learning has emerged from the disciplines of knowledge engineering, which is known as *inductive logic programming* [Konar, 2000].

Supervised learning: In supervised learning a trainer submits the input–output exemplary patterns and the learner has to adjust the parameters of the system autonomously, so that it can yield the correct output pattern when excited with one of the given input patterns. Inductive learning [Michalski, 1983] is a special class of the supervised learning technique, where, given a set of $\{x_i, f(x_i)\}$ pairs, a hypothesis $h(x_i)$ is determined such that $h(x_i) \approx f(x_i), \forall i$. This demonstrates that a number of training instances are required to form a concept in inductive learning.

Unsupervised learning: Unsupervised learning employs no trainer and the learner has to construct concepts by experimenting on the environment. The environment responds but fails to identify which ones are rewarding and which ones are punishable activities. This is because of the fact that the goals or the outputs of the training instances are unknown. Therefore, the environment cannot measure the status of the activities of the learner with respect to the goals. One of the simplest ways to construct a concept by unsupervised learning is through experiments. For example, suppose a child throws a ball to the wall; the ball bounces and returns to the child. After performing this experiment a number of times, the child learns the 'principle of bouncing', which is an example of unsupervised learning.

An intelligent system should be able to learn quickly from large amounts of data [Kasabov, 1998]. It is also stated that the machine should adapt in real time and in an online mode as new data is encountered. It should be memory based and possess data and exemplary storage and retrieval capacities. Secondly, the system should be able to learn and improve through active interaction with the user and the environment. But not much progress has been achieved for this learning to date.

Reinforcement learning: In reinforcement learning, the learner adapts its parameter by determining the status, i.e. reward or punishment of the feedback signals from its environment. The simplest form of reinforcement learning is adopted in learning automata. Currently Q learning and temporal difference learning have been devised based on the reward/punishment status of the feedback signals.

1.5 Soft Computing Tools and Robot Cognition

A collection of tools shared by *artificial neural nets, fuzzy logic, genetic algorithms*, *belief calculus*, and some aspects of *inductive logic programming* are known as *soft computing tools*. These tools are used independently as well as jointly depending on the type of the domain of application [Jain, 1999]. According to Zadeh, soft computing is "an emerging approach for computing, which parallels the remarkable ability of the human mind to reason and learn in an environment of uncertainty and imprecision" [Zadeh, 1983]. The scope of these tools in modeling robot cognition is outlined below.

1.5.1 Modeling Cognition Using ANN

As we know the goal of cognitive modeling is the development of algorithms that require machines to perform cognitive tasks at which humans are presently better [Haykins, 1999]. A cognitive system must be capable of (i) sensing the external environment and storing it in the form of knowledge, (ii) applying the knowledge stored to solve problems, and (iii) acquiring new knowledge through experience. To perform this task the machine needs representation, reasoning and learning.

Machine learning may involve two different kinds of information processing, i.e. inductive and deductive. In inductive processing, generally patterns and rules are determined from raw data and experience; whereas in deductive processing general rules are used to determine specific facts. Similarity-based learning uses induction whereas the proof of a theorem is a deduction from known axioms and other existing theorems. Both induction and deduction processing can be used for explanation-based learning. The importance of knowledge bases and difficulties experienced in learning has led to the development of various methods for supplementing knowledge bases. Specifically, if there are experts in a given field, it is usually easier to obtain the compiled experience of the experts, rather than use direct experience. In fact, this is the idea behind the development of neural networks as a cognitive model. Let us compare neural networks with the cognitive model with respect to three aspects, namely level of explanation, style of processing and representation of structure.

Level of explanation: In traditional machine intelligence, the emphasis is given to building symbolic representations of the problem. It assumes the existence of a mental representation and it models cognition as the sequential processing of symbolic representations [Newell et al., 1972]. On the other hand, a neural network emphasizes the development of parallel distribution processing models. These models assume that information processing takes place through the interaction of a large number of neurons, each of which sends excitatory and inhibitory signals to other neurons in the network [Rumelhart et al., 1986]. Moreover, neural networks give more emphasis to the neurobiological explanation of cognitive phenomenon.

Style of processing: In traditional machine intelligence, processing is sequential as in typical computer programming. Even when there is no predetermined order, the operations are performed in a stepwise manner. On the other hand, neural networks process the information in parallel and provide flexibility about the structure of the source. Moreover, parallelism may be massive which gives neural networks a remarkable form of robustness. With the computation spread over many neurons, it usually does not matter much if the states of some neurons in the network deviate from their expected values. Noisy or incomplete inputs may still be recognized, and may be able to function satisfactorily and therefore learning does not have to be perfect. Performance of the network degrades within a certain range. The network is made even more robust by virtue of coarse coding, where each feature is spread over several neurons [Hinton, 1981].

Representation of structure: In traditional machine intelligence, representation is done through a language of thought, which possesses a quasi-linguistic structure. These are generally complex to build in a systematic fashion from simple symbols. In contrast, the nature and structure of representation is very crucial in neural networks. For implementation of cognitive tasks, a neural network emphasizes the approach to building a structure connectionist model that integrates them. As a result a neural network combines the desirable features of adaptability, robustness and uniformity with representation and inference [Feldman, 1992; Waltz, 1997].

The ANNs adjust the weights of the neurons between different layers during the adaptation cycle. The adaptation cycle is required for updating various parameters of the network, until a state of equilibrium is reached, following which the parameters no longer change. ANNs support both supervised and unsupervised learning as mentioned earlier. The supervised learning algorithms realized with ANN have been successfully applied in control, automation, robotics and computer vision [Narendra et al., 1990]. On the other hand, unsupervised learning algorithms built with ANNs have been applied in *scheduling, knowledge acquisition* [Buchanan, 1993], *planning* [McDermott et al., 1984] and *analog to digital conversion of data* [Sympson, 1988].

1.5.2 Fuzzy Logic in Robot Cognition

Fuzzy logic deals with fuzzy sets and logical connectives for modeling the human-like reasoning problems of the real world. A fuzzy set, unlike conventional sets, includes all elements of the universal set of the domain with varying membership values in the interval [0,1]. It may be noted that a conventional set contains its members with a membership value equal to one and disregards other elements of the universal set with a zero membership value. The most common operators applied to fuzzy sets are AND (minimum), OR (maximum) and negation (complementation), where AND and OR have binary arguments, while negation has a unary argument. The logic of *Fuzzy Set Theory* was proposed by Zadeh [Zadeh, 1983], who introduced the concept of system theory, and later extended it for approximate reasoning in expert systems. Other pioneering research contributions on *Fuzzy Logic* include the work of Tanaka in stability analysis of control systems [Tanaka, 1995], Mamdani in *cement kiln control* [Mamdani, 1977], Kosko [Kosko, 1994] and Pedrycz [Pedrycz, 1995] in *Fuzzy Neural Nets*, Bezdek in *Pattern Classification* [Bezdek, 1991], and Zimmerman [Zimmerman, 1991] and Yager [Yager, 1983] in *Fuzzy Tools and Techniques*.

Fuzzy logic has become a popular tool for robot cognition in recent years [Saffioti, 1997]. Given the uncertain and incomplete information about the environment available to the autonomous robot, fuzzy rules provide an attractive means for mapping ambiguous sensor data to appropriate information in real time. The methodology of fuzzy logic appears very useful when the processes are too complex for analysis by conventional quantitative techniques or when the available sources of information are interpreted qualitatively, inexactly, or uncertainly, which is the case with mobile robots. However, fuzzy logic parameters are usually determined by domain experts using a trial and error method. Also, as the number of input variables increases, in the case of mobile robots, the number of rules increases exponentially, and this creates much difficulty in determining a large number of rules.

1.5.3 Genetic Algorithms in Robot Cognition

Genetic algorithms (GAs) are stochastic in nature, and mimic the natural process of biological evolution [Rich et al., 1996]. This algorithm borrows the principle of *Darwinism*, which rests on the fundamental belief of the *survival of the fittest* in the process of natural selection of species. GAs find extensive applications in intelligent search, machine learning and optimization problems. The problem states in a GA are denoted by chromosomes, which are usually represented by binary strings. The most common operators used in GAs are crossover and mutation. The evolutionary cycle in a GA consists of the following three sequential steps [Michalewicz, 1986];

(i) generation of a population (problem states represented by chromosomes)

(ii) selection of better candidate states from the generated population

(iii) genetic evolution through crossover followed by mutation.

In step (i) a few initial problem states are first identified and in step (ii) a fixed number of better candidate states are selected from the generated population. Step (iii) evolves a new generation through the process of crossover and mutation. These steps are repeated a finite number of times to obtain the solution for the given problem.

GAs have been successfully applied to solve a variety of theoretical and practical problems by imitating the underlying processes of evolution, such as selection, recombination, and mutation. The GA-based approach is a well-accepted technique for enabling systems to adapt to different control tasks [Filho, 1994]. But, it is not feasible for a simple GA to learn online and adapt in real time. The situation is worsened by the fact that most GA methods developed so far assume that the solution space is fixed, thus preventing them from being used in real-time applications [Michalewicz, 1986].

1.6 Summary

This chapter briefly highlights the development of various models of mobile robots and the paradigm shift towards the model of cognition. A model of cognition has been introduced here, which will be realized for various tasks of simulated robots as well as for the mobile robots in subsequent chapters. The chapter defines the term ‘cognition’ along with its embedded cycles namely the acquisition cycle, perception cycle, learning and coordination cycle and their associated states. A brief review of visual perception, visual recognition and machine learning has been given in subsequent sections. The application of various soft computing tools like fuzzy logic, genetic algorithms and artificial neural networks for robot cognition has also been outlined.

Springer

http://www.springer.com/978-3-540-23446-3

Robot Cognition and Navigation
An Experiment with Mobile Robots
Patnaik, S.
2007, XV, 291 p., Hardcover
ISBN: 978-3-540-23446-3