CHAPTER 1

Introduction to Reinforcement and Systemic Machine Learning

1.1 INTRODUCTION

The expectations from intelligent systems are increasing day by day. What an intelligent system was supposed to do a decade ago is now expected from an ordinary system. Whether it is a washing machine or a health care system, we expect it to be more and more intelligent and demonstrate that behavior while solving complex as well as day-to-day problems. The applications are not limited to a particular domain and are literally distributed across all domains. Hence domain-specific intelligence is fine but the user has become demanding, and a true intelligent and problem-solving system irrespective of domains has become a necessary goal. We want the systems to drive cars, play games, train players, retrieve information, and help even in complex medical diagnosis. All these applications are beyond the scope of isolated systems and traditional preprogrammed learning. These activities need dynamic intelligence. Dynamic intelligence can be exhibited through learning not only based on available knowledge but also based on the exploration of knowledge through interactions with the environment. The use of existing knowledge, learning based on dynamic facts, and acting in the best way in complex scenarios are some of the expected features of intelligent systems.

The learning has many facets. Right from simple memorization of facts to complex inference are some examples of learning. But at any point of time, learning is a holistic activity and takes place around the objective of better decision-making. Learning results from data storing, sorting, mapping, and classification. Still one of the most important aspects of intelligence is learning. In most of the cases we expect learning to be a more goal-centric activity. Learning results from an inputs from an experienced person, one's own experience, and inference based on experiences or past learning. So there are three ways of learning:

- Learning based on expert inputs (supervised learning)

Reinforcement and Systemic Machine Learning for Decision Making, First Edition. Parag Kulkarni.

Published 2012 by John Wiley & Sons, Inc.

- Learning based on own experience
- Learning based on already learned facts

In this chapter, we will discuss the basics of reinforcement learning and its history. We will also look closely at the need of reinforcement learning. This chapter will discuss limitations of reinforcement learning and the concept of systemic learning. The systemic machine-learning paradigm is discussed along with various concepts and techniques. The chapter also covers an introduction to traditional learning methods. The relationship among different learning methods with reference to systemic machine learning is elaborated in this chapter. The chapter builds the background for systemic machine learning.

1.2 SUPERVISED, UNSUPERVISED, AND SEMISUPERVISED MACHINE LEARNING

Learning that takes place based on a class of examples is referred to as *supervised learning*. It is learning based on labeled data. In short, while learning, the system has knowledge of a set of labeled data. This is one of the most common and frequently used learning methods. Let us begin by considering the simplest machine-learning task: *supervised learning* for classification. Let us take an example of classification of documents. In this particular case a learner learns based on the available documents and their classes. This is also referred to as labeled data. The program that can map the input documents to appropriate classes is called a *classifier*, because it assigns a class (i.e., document type) to an object (i.e., a document). The task of supervised learning is to construct a classifier given a set of classified training examples. A typical classification is depicted in Figure 1.1.

Figure 1.1 represents a hyperplane that has been generated after learning, separating two classes—class A and class B in different parts. Each input point presents input–output instance from sample space. In case of document classification, these points are documents. Learning computes a separating line or hyperplane among documents. An unknown document type will be decided by its position with respect to a separator.

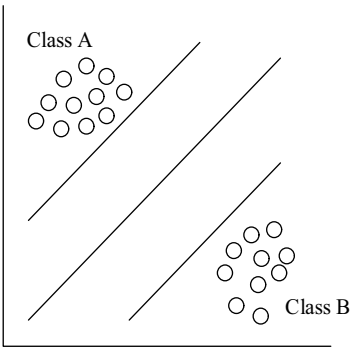

Figure 1.1 Supervised learning.

There are a number of challenges in supervised classification such as generalization, selection of right data for learning, and dealing with variations. *Labeled examples* are used for training in case of supervised learning. The set of labeled examples provided to the learning algorithm is called the *training set*.

The classifier and of course the decision-making engine should minimize false positives and false negatives. Here false positives stand for the result yes—that is, classified in a particular group wrongly. False negative is the case where it should have been accepted as a class but got rejected. For example, apples not classified as apples is false negative, while an orange or some other fruit classified as an apple is false positive in the apple class. Another example of it is when guilty but not convicted is false positive, while innocent but convicted or declared innocent is false negative. Typically, wrongly classified are more harmful than unclassified elements.

If a classifier knew that the data consisted of sets or batches, it could achieve higher accuracy by trying to identify the boundary between two adjacent sets. It is true in the case of sets of documents to be separated from one another. Though it depends on the scenario, typically false negatives are more costly than false positives, so we might want the learning algorithm to prefer classifiers that make fewer false negative errors, even if they make more false positives as a result. This is so because false negative generally takes away the identity of the objects or elements that are classified correctly. It is believed that the false positive can be corrected in next pass, but there is no such scope for false negative.

Supervised learning is not just about classification, but it is the overall process that with guidelines maps to the most appropriate decision.

Unsupervised learning refers to learning from unlabeled data. It is based more on similarity and differences than on anything else. In this type of learning, all similar items are clustered together in a particular class where the label of a class is not known.

It is not possible to learn in a supervised way in the absence of properly labeled data. In these scenarios there is need to learn in an unsupervised way. Here the learning is based more on similarities and differences that are visible. These differences and similarities are mathematically represented in unsupervised learning.

Given a large collection of objects, we often want to be able to understand these objects and visualize their relationships. For an example based on similarities, a kid can separate birds from other animals. It may use some property or similarity while separating, such as the birds have wings. The criterion in initial stages is the most visible aspects of those objects. Linnaeus devoted much of his life to arranging living organisms into a hierarchy of classes, with the goal of arranging similar organisms together at all levels of the hierarchy. Many unsupervised learning algorithms create similar hierarchical arrangements based on similarity-based mappings. The task of *hierarchical clustering* is to arrange a set of objects into a hierarchy such that similar objects are grouped together. *Nonhierarchical clustering* seeks to partition the data into some number of disjoint clusters. The process of clustering is depicted in Figure 1.2. A learner is fed with a set of scattered points, and it generates two clusters with representative centroids after learning. Clusters show that points with similar properties and closeness are grouped together.

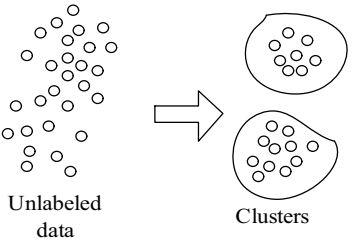

Figure 1.2 Unsupervised learning.

In practical scenarios there is always need to learn from both labeled and unlabeled data. Even while learning in an unsupervised way, there is the need to make the best use of labeled data available. This is referred to as *semisupervised learning*. Semisupervised learning is making the best use of two paradigms of learning—that is, learning based on similarity and learning based on inputs from a teacher. Semisupervised learning tries to get the best of both the worlds.

1.3 TRADITIONAL LEARNING METHODS AND HISTORY OF MACHINE LEARNING

Learning is not just knowledge acquisition but rather a combination of knowledge acquisition, knowledge augmentation, and knowledge management. Furthermore, intelligent inference is essential for proper learning. Knowledge deals with significance of information and learning deals with building knowledge. How can a machine can be made to learn? This research question has been posed for more than six decades by researchers. The outcome of this research has built a platform for this chapter. Learning involves every activity. One such example, is the following: While going to the office yesterday, Ram found road repair work in progress on route one, so he followed route two today. It might be possible that route two is worse. Then he may go back to route one or might try route three. Route one is in bad shape due to repair work is knowledge built, and based on that knowledge he has taken action: following route 2, that is, exploration. The complexity of learning increases as the number of parameters and time dimensions start playing a role in decision making.

Ram found that road repair work is in progress on route one.

He hears an announcement that in case of rain, route two will be closed.

He needs to visit a shop X while going to office.

He is running out of petrol.

These new parameters make his decision much more complex as compared to scenario 1 and scenario 2 discussed above.

In this chapter, we will discuss various learning methods along with examples. The data and information used for learning are very important. The data cannot be

used as is for learning. It may contain outliers and information about features that may not be relevant with respect to the problem one is trying to solve. The approaches for the selection of data for learning vary with the problems. In some cases the most frequent patterns are used for learning. Even in some cases, outliers are also used for learning. There can be learning based on exceptions. The learning can take place based on similarities as well as differences. The positive as well as negative examples help in effective learning. Various models are built for learning with the objective of exploiting the knowledge.

Learning is a continuous process. The new scenarios are observed and new situations arise—those need to be used for learning. Learning from observation needs to construct meaningful classification of observed objects and situation. Methods of measuring similarity and proximity are employed for this purpose. Learning from observations is the most commonly used method by human beings. While making decisions we may come across the scenarios and objects that we have not used or came across during a learning phase. The inference allows us to handle these scenarios. Furthermore, we need to learn in different and new scenarios and hence even while making decisions the learning continues.

There are three fundamental continuously active human-like learning mechanisms:

1. *Perceptual Learning:* Learning of new objects, categories, and relations. It is more like constantly seeking to improve and grow. It is similar to the learning professionals use.
2. *Episodic Learning:* It is based on events and information about the event, like what, where, and when. It is the learning or the change in the behavior that occurs due to an event.
3. *Procedural Learning:* Learning based on actions and action sequences to accomplish a task. Implementation of this human cognition can impart intelligence to a machine. Hence, a unified methodology around intelligent behavior is the need of time that will allow machines to learn and behave or respond intelligently in dynamic scenarios.

Traditional machine-learning approaches are susceptible to dynamic continual changes in the environment. However, perceptual learning in human does not have such restrictions. Learning in humans is selectively incremental, so it does not need a large training set and is simultaneously not biased by already learned but outdated facts. Learning and knowledge extraction in human beings is dynamic, and a human brain adapts to changes occurring in the environment continuously.

Interestingly, psychologists have played a major role in the development of machine-learning techniques. It has been a movement taken by computer researchers and psychologists together to make machines intelligent for more than six decades. The application areas are growing, and research done in the last six decades made us believe that it is one of the most interesting areas to make machines learn.

Machine learning is the study of methods for programming computers to learn. It is about making machines to behave intelligently and learn from experiences like human beings. In some tasks the human expert may not be required; this may include automated manufacturing or repetitive tasks with very few dynamic situations but demanding very high level of precision. A machine-learning system can study recorded data and subsequent machine failures and learn prediction rules. Second, there are problems where human experts exist and are required, but the knowledge is present in a tacit form. Speech recognition and language understanding come under this category. Virtually all humans exhibit expert-level abilities on these tasks, but the exact method and steps to perform these tasks are not known. A set of inputs and outputs with mapping is provided in this case, and thus machine-learning algorithms can learn to map the inputs to the outputs.

Third, there are problems where phenomena are changing rapidly. In real life there are many dynamic scenarios. Here the situations and parameters are changing dynamically. These behaviors change frequently, so that even if a programmer could construct a good predictive computer program, it would need to be rewritten frequently. A learning program can relieve the programmer of this burden by constantly modifying and tuning a set of learned prediction rules.

Fourth, there are applications that need to be customized for each computer user separately. A machine-learning system can learn the customer-specific requirements and tune the parameters accordingly to get a customized version for a specific customer.

Machine learning addresses many of the research questions with the aid of statistics, data mining, and psychology. Machine learning is much more than just data mining and statistics. Machine learning (ML) as it stands today is the use of data mining and statistics for inferencing to make decisions or build knowledge to enable better decision making. Statistics is more about understanding data and the pattern between them. Data mining seeks the relevant data based on patterns for decision making and analysis. Psychological studies of human learning aspire to understand the mechanisms underlying the various learning behaviors exhibited by people. At the end of the day, we want machine learning to empower machines with the learning abilities that are demonstrated by humans in complex scenarios. The psychological studies of human nature and the intelligence also contribute to different methods of machine learning. This includes concept learning, skill acquisition, strategy change, analytical inferences, and bias based on scenarios.

Machine learning is primarily concerned with the timely response, accuracy, and effectiveness of the resulting computer system. It many times does not take into account other aspects such as learning abilities and responding to dynamic situations, which are equally important. A machine-learning approach focuses on many complex applications such as building an accurate face recognition and authentication system. Statisticians, psychologists, and computer scientists may work together on this front. A data mining approach might look for patterns and variations in image data.

One of the major aspects of learning is the selection of learning data. All the information available for learning cannot be used as it is. It may contain a lot of data

that may not be relevant or captured from a completely different perspective. Every bit of data cannot be used with the same importance and priority. The prioritization of the data is done based on scenarios, system significance, and relevance. The determination of relevance of these data is one of the most difficult parts of the process.

There are a number of challenges in making machines learn and making suitable decisions at the right time. The challenges start from the availability of limited learning data, unknown perspectives, and defining the decision problems. Let us take a simple example where a machine is expected to prescribe the right medicine to a patient. The learning set may include samples of patients, their histories, their test reports, and the symptoms reported by them. Furthermore, the data for learning may also include other information such as family history, habits, and so on. In case of a new patient, there is the need to infer based on available limited information because the manifestation of the same disease may be different in his case. Some key information might be missing, and hence decision making may become even more difficult.

When we look at the way a human being learns, we find many interesting aspects. Generally the learning takes place with understanding. It is facilitated when new and existing knowledge is structured around the major concepts and principles of the discipline. During the learning, either some principles are already there or developed in the process work as a guideline for learning. The learning also needs prior knowledge. Learners use what they already know to construct new understandings. This is more like building knowledge. Furthermore, there are different perspectives and metacognition. Learning is facilitated through the use of metacognitive strategies that identify, monitor, and regulate cognitive processes.

1.4 WHAT IS MACHINE LEARNING?

A general concept of machine learning is depicted in Figure 1.3. Machine learning studies computer algorithms for learning. We might, for instance, be interested in learning to complete a task, or to make accurate predictions, reactions in certain situations, or to behave intelligently. The learning that is being done is always based on some sort of observations or data, such as examples (the most common case in this course), direct experience, or instruction. So in general, machine learning is about learning to do better in the future based on what was experienced in the past. It is making a machine to learn from available information, experience, and knowledge building.

In the context of the present research, machine learning is the development of programs that allow us to analyze data from the various sources, select relevant data,

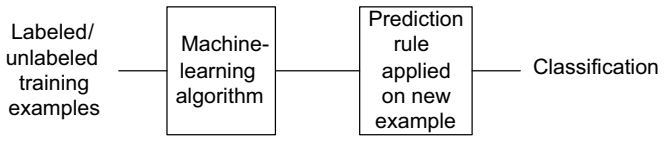

Figure 1.3 Machine learning and classification.

and use those data to predict the behavior of the system in another similar and if possible different scenario. Machine learning also classifies objects and behaviors to finally impart the decisions for new input scenarios. The interesting part is that more learning and intelligence is required to deal with uncertain situations.

1.5 MACHINE-LEARNING PROBLEM

It can be easily concluded that all the problems that need intelligence to solve come under the category of machine-learning problems. Typical problems are character recognition, face authentication, document classification, spam filtering, speech recognition, fraud detection, weather forecasting, and occupancy forecasting. Interestingly, many problems that are more complex and involve decision making can be considered as machine-learning problems as well. These problems typically involve learning from experiences and data, and search for the solutions in known as well as unknown search spaces. It may involve the classification of objects, problems, and mapping them to solutions or decisions. Even classification of any type of objects or events is also a machine-learning problem.

1.5.1 Goals of Learning

The primary goal of learning/machine learning is producing some learning algorithm with practical value. In the literature and research, most of the time machine learning is referred to from the perspective of applications and it is more bound by methods. The goals of ML are described as development and enhancement of computer algorithms and models to meet the decision-making requirements in practical scenarios. Interestingly, it did achieve the set goal in many applications. Right from washing machines and microwave ovens to the automated landing of aircraft, machine learning is playing a major role in all modern applications and appliances. The era of machine learning has introduced methods from simple data analysis and pattern matching to fuzzy logic and inferencing.

In machine learning, most of the inferencing is data driven. The sources of data are limited and many times there is difficulty in identifying the useful data. It may be possible that the source contains large piles of data and that the data contain important relationships and correlations among them. Machine learning can extract these relationships, which is an area of data mining applications. The goal of machine learning is to facilitate in building intelligent systems (IS) that can be used in solving real-life problems.

The computational power of the computing engine, the sophistication and elegance of algorithms, the amount and quality of information and values, and the efficiency and reliability of the system architecture determine the amount of intelligence. The amount of intelligence can grow through algorithm development, learning, and evolution. Intelligence is the product of natural selection, wherein more successful behavior is passed on to succeeding generations of intelligent systems and less successful behavior dies out. This intelligence helps humans and intelligent systems to learn.

In supervised learning we learn from different scenarios and expected outcomes presented as a learning material. The purpose is that if we come across a similar scenario in the future we should be in position to make appropriate or rather the best possible decisions. This is possible if we can classify a new scenario to one of the known classes or known scenarios. Enabling to classify the new scenario allows us to select an appropriate action. Learning is possible by imitation, memorization, mapping, and inference. Furthermore, induction, deduction, and example-based and observation-based learning are some other ways in which learning is possible.

Learning is driven by objective and governed by certain performance elements and their components. The clarity about the performance elements and their components, available feedback to learn the behavior of these components, and the representation of these components are necessary for learning. The agents need to learn, and components of these agents should be able to map and determine actions, extract and infer about the information related to the environment, and set goals that describe classes of states. The desired actions with reference to value or state help the system to learn. The learning takes place based on feedbacks. These feedbacks come in the form of penalties or rewards.

1.6 LEARNING PARADIGMS

An empirical learning method has three different approaches to modeling problems based on observation, data, and partial knowledge about problem domains. These approaches are more specific to problem domains. They are

1. Generative modeling
2. Discriminative modeling
3. Imitative modeling

Each of these models has their own pros and cons. They are best suited for different application areas depending on training samples and prior knowledge. Generally, learning model suitability depends on the problem scenario and available knowledge and decision complexities.

In a *generative modeling approach*, statistics provide a formal method for determining nondeterministic models by estimating joint probability over variables of problem domain. Bayesian networks are used to capture dependencies among domain variables as well as distributions among them. This partial domain knowledge combined with observations enhances the probability density function. Generative density function is then used to generate samples of different configurations of the system and to draw an inference on an unknown situation. Traditional rule-based expert systems are giving way to statistical generative approaches due to visualization of interdependencies among variables that yields better prediction than heuristic approaches. Natural language processing, speech recognition, and topic modeling among different speakers are some of the application areas of generative modeling. This probabilistic approach of learning can be used in computer vision, motion

tracking, object recognition, face recognition, and so on. In a nutshell, learning with generative modeling can be applied in the domains of perception, temporal modeling, and autonomous agents. This model tries to represent and model interdependencies in order to lead to better predictions.

A discriminative approach models posterior probability or discriminant functions with less domain-specific or prior knowledge. This technique directly optimizes target task-related criteria. For example, a Support Vector Machine maximizes the margin of a hyperplane between two sets of variables in *n* dimensions. This approach can be widely used for document classification, character recognition, and other numerous areas where interdependency among problem variables does not play any role or play the minimum role in observation variables. Thus, prediction is not influenced by inherent problem structure and also by domain knowledge. This approach may not be very effective in the case of very high level of interdependencies.

The third approach is *imitative learning*. Autonomous agents, which exhibit interactive behavior, are trained through an imitative learning model. The objective of imitative leaning is to learn an agent's behavior by providing a real example of agents' interaction with the world and generalizing it. The two components of this learning model, passively perceiving real-world behavior and learning from it, are depicted in Figure 1.4. Interactive agents perceive the environment using a generative model to regenerate/synthesize virtual characters/interaction and use a discriminative approach on temporal learning to focus on the prediction task necessary for action selection. An agent tries to imitate real-world situations with intelligence so that if exact behavior is not available in a learned hypothesis, the agent can still take some action based on synthesis. Occurrence of imitative and observational learning can be used in contingence with reinforcement learning. The imitative response can be the action for the rewards in reinforcement learning.

Figure 1.4 depicts the imitative learning with reference to a demonstrator and environment. The demonstration is rather action or a series of actions from which an observer learns. Environment refers to the environment of the observer. The learning takes place based on imitation and observation of demonstration while knowledge base and environment help in inferring different facts to complete the learning. Imitative learning can be extended to imitative reinforcement learning where imitation is based on previous knowledge learning and the rewards are compared with pure imitative response.

Learning based on experience need to have input and outcome of experience to be measured. For any action there is some outcome. The outcome leads to some sort of amendment in your action. Learning can be data-based, event-based, pattern-based, and system-based. There are advantages and disadvantages of each of these paradigms of learning. Knowledge building and learning is a continuous process, and we would like systems to reuse creatively and intelligently what is learned selectively in order to achieve the goal state.

Interestingly, when a kid is learning to walk, it is using all types of learning simultaneously. It has some supervised learning in the form of parents guiding it, some unsupervised learning based on new data points it is coming across, inference for

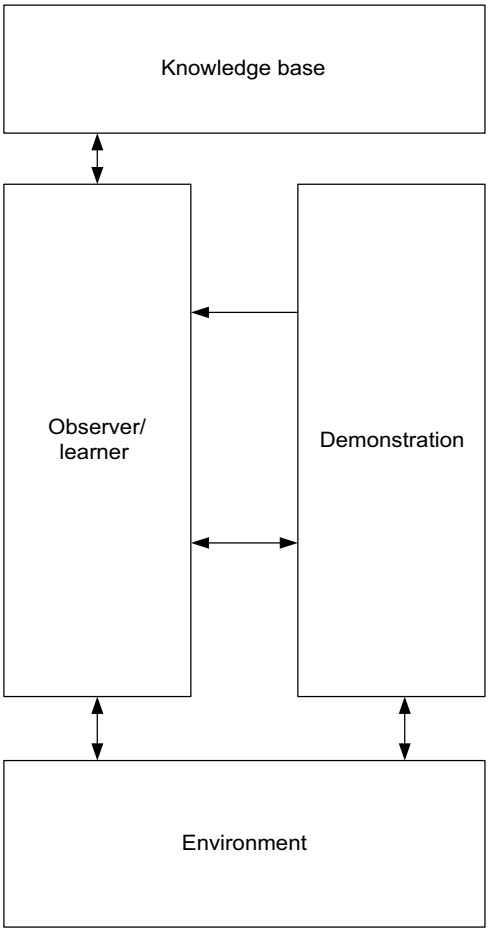

Figure 1.4 Reinforcement and imitative learning.

some similar scenarios, and feedback from environment. Learning results from labeled as well as unlabeled data, and it takes place simultaneously. In fact a kid is using all the learning methods and much more than that. A kid not only uses available knowledge and context but also infers information that cannot be derived directly from the available data. Kids use all these methods selectively, together, and based on need and appropriateness. The learning by kids results from their close interactions with environment. While making systems learn from experiences, we need to take into account all these facts. Furthermore, it is more about paradigms rather than methods used for learning. This book is about making a system intelligent with focus on reinforcement learning. Reinforcement learning tries to strike balance between exploitation and exploration. Furthermore, it takes place with interaction with environment. Rewards from environment and then cumulative value drive the overall actions. Figure 1.5 depicts the process of how a kid learns. Kids get many

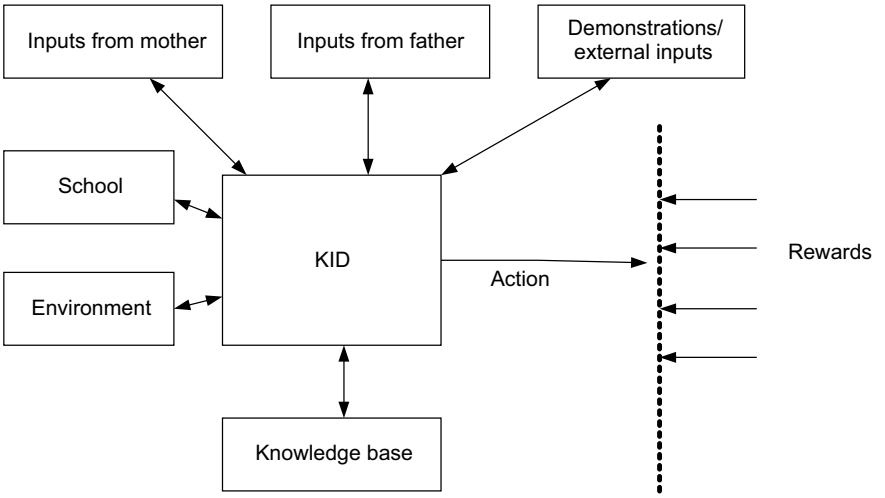

Figure 1.5 Kid learning model.

inputs from their parents, society, school, and experiences. They perform actions, and for actions they obtain rewards from these sources and environment.

1.7 MACHINE-LEARNING TECHNIQUES AND PARADIGMS

The learning paradigm kept changing over the years. The concept of intelligence changed, and even paradigm of learning and knowledge acquisition changed. Paradigm is (in the philosophy of science) a very general conception of the nature of scientific endeavor within which a given enquiry is undertaken. The learning as per Peter Senge is the acquiring of information and knowledge that can empower us to get what we would like to get out of life [1].

In machine learning if we go through the history, learning is initially assumed more as memorization and getting or reproducing one of the memorized facts that is appropriate when required. This paradigm can be called a data-centric paradigm. In fact this paradigm does exist in machine learning even today and is being used to great extent in all intelligent programs. Take the example of a simple program of retrieving the age of employees. A simple database with names and age is maintained; and when the name of any employee is given, the program can retrieve the age of the given employee. There are many such database-centric applications demonstrating data centric intelligence. But slowly the expectations from intelligent systems started increasing. As per the Turing test of intelligence, an intelligent system is one that can behave like a human, or it is difficult to make out whether a response is coming from a machine or a human.

The learning is interdisciplinary and deals with many aspects from psychology, statistics, mathematics, and neurology. Interestingly, all human behaviors could not correspond to intelligence and hence there are some areas where a computer can

behave or respond in a better way. The Turing test is applicable to intelligent behavior of computers. There are some intelligent activities that humans do not do or that, machines can do in a better way than humans.

Reinforcement learning is making systems get the best of both worlds in the best possible way. But since the systemic behaviors of activities and decision making makes it necessary to understand the system behavior and components for effective decision making, traditional paradigms of machine learning may not exhibit the required intelligent behavior in complex systems. Every activity, action, and decision has some systemic impact. Furthermore, any event may result from some other event or series of events from a systemic perspective. These relationships are complex and difficult to understand. Systemic machine learning is more about exploitation, exploration from systemic perspective to build knowledge to get what we expect from the system. Learning from experience is the most important part of it. With more and more experience the behavior is expected to improve.

Two aspects of learning include learning for predictable environment behavior and learning for nonpredictable environment behavior. As we expect systems and machines to behave intelligently even in a nonpredictive environment, we need to look at learning paradigms and models from the perspective of new expectations. These expectations make it necessary to learn continuously and from various sources of information.

Representing and adapting knowledge for these systems and using them effectively is a necessary part of it. Another important aspect of learning is context: the intelligence and decision making should make effective use of context. In the absence of context, deriving the meaning of data is difficult. Further decisions may differ as per the context. Context is very systemic in nature. Context talks more about the scenario—that is, circumstances and facts surrounding the event. In absence of the facts and circumstances of related data, decision making becomes a difficult task. The context covers various aspects of environment and system such as environmental parameters, interactions with other systems and subsystems, various parameters, and so on. A doctor asks patients a number of questions. The information given by a patient along with the information with the doctor about epidemic and other recent health issues and outcome of conducted medical tests builds context for him/her. A doctor uses this context to diagnose.

The intelligence is not isolated and needs to use information from the environment for decision making as well as learning. The learning agents get feedback in the form of reward/penalty for their every action. They are supposed to learn from experience. To learn, there is a need to acquire more and more information. In real-life scenarios the agents cannot view anything and everything. There are fully observable environments and partially observable environments. Practically all environments are partially observable unless specific constraints are posed for some focused goal. The limited view limits the learning and decision-making abilities. The concept of integrating information is used very effectively in intelligent systems—the learning paradigm is confined by data-centric approaches. The context considered in the past research was more data centric and was never at a center of the activity.

1.8 WHAT IS REINFORCEMENT LEARNING?

There are tons of nonlinear and complex problems still waiting for the solutions. Ranging from automated car drivers to next level security systems. These problems look solvable—but the methods, solutions, and available information are just not enough to provide a graceful solution.

The main objective in solving a machine-learning problem is to produce intelligent programs or intelligent agents through the process of learning and adapting to changed environment. Reinforcement learning is one such machine-learning process. In this approach, learners or software agents learn from direct interaction with environment. This mimics the way human being learns. The agent can also learn even if complete model or information about environment is not available. An agent gets feedback about its actions as reward or punishment. During a learning process, these situations are mapped to actions in an environment. Reinforcement learning algorithms maximize rewards received during interactions with environment and establish the mapping of states to actions as a decision-making policy. The policy can be decided once or it can also adapt with changes in environment.

Reinforcement learning is different from *supervised learning*—the most widely used kind of learning. Supervised learning is learning from examples provided by a knowledgeable external supervisor. It is a method for training a parameterized function approximator. But it is not adequate for learning from interaction. It is more like learning from external guidance, and the guidance sits out of the environment or situation. In interactive problems it is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act. In uncharted territory, where one would expect learning to be most beneficial, an agent must be able to learn from its own experience and from environment also. Thus, reinforcement learning combines the field of dynamic programming and supervised learning to generate a machine-learning system, which is very close to approaches used by human learning.

One of the challenges that arise in reinforcement learning and not in other kinds of learning is the trade-off between exploration and exploitation. To obtain a lot of reward, a reinforcement-learning agent must prefer actions that it has tried in the past and found to be effective in producing reward. But to discover such actions, it has to try actions that it has not selected before. The agent has to *exploit* what it already knows in order to obtain reward, but it also has to *explore* in order to make better action selections in the future. The dilemma is that neither exploration nor exploitation can be pursued exclusively without failing at the task. The agent must try a variety of actions *and* progressively favor those that appear to be best. On a stochastic task, each action must be tried many times to gain a reliable estimate of its expected reward. The entire issue of balancing exploration and exploitation does not arise in supervised learning, as it is usually defined. Furthermore, supervised learning never looks into exploration, and the responsibility of exploration is given to experts.

Another key feature of reinforcement learning is that it explicitly considers the *whole* problem of a goal-directed agent interacting with an uncertain environment.

This is in contrast with many approaches that consider subproblems without addressing how they might fit into a larger picture. For example, we have mentioned that much of machine-learning research is concerned with supervised learning without explicitly specifying how such ability would finally be useful. Other researchers have developed theories of planning with general goals, but without considering planning's role in real-time decision making nor considering the question of where the predictive models necessary for planning would come from. Although these approaches have yielded many useful results, their focus on isolated subproblems is a significant limitation. These limitations come from the inability to interact in real-time scenarios and the absence of active learning.

Reinforcement learning differs from the more widely studied problem of supervised learning in several ways. The most important difference is that there is no presentation of input–output pairs. Instead, after choosing an action the agent is told the immediate reward and the subsequent state, but is *not* told which action would have been in its best long-term interests. It is necessary for the agent to gather useful experience about the possible system states, actions, transitions, and rewards actively to act optimally. Another difference from supervised learning is that online performance is important; the evaluation of the system is often concurrent with learning.

Reinforcement learning takes the opposite track, starting with a complete, interactive, goal-seeking agent. All reinforcement-learning agents have explicit goals, can sense aspects of their environments, and can choose actions to influence their environments. Moreover, it is usually assumed from the beginning that the agent has to operate despite significant uncertainty about the environment it faces. When reinforcement learning involves planning, it has to address the interplay between planning and real-time action selection, as well as the question of how environmental models are acquired and improved. When reinforcement learning involves supervised learning, it does so for specific reasons that determine which capabilities are critical and which are not.

Some aspects of reinforcement learning are closely related to search and planning issues in artificial intelligence (AI), especially in the case of intelligent agents. AI search algorithms generate a satisfactory trajectory through a graph of states. The search algorithms are focused on searching a goal state based on informed or uninformed methods. The combination of informed and uninformed methods is similar to exploration and exploitation of knowledge. Planning operates in a similar manner, but typically within a construct with more complexity than a graph, in which states are represented by compositions of logical expressions instead of atomic symbols. These AI algorithms are less general than the reinforcement-learning methods, where the AI algorithms require a predefined model of state transitions and with a few exceptions assumed. These methods are typically confined by predefined models and well-defined constraints. On the other hand, reinforcement learning, at least in the form of discrete cases, assumes that the entire state space can be enumerated and stored in memory—an assumption to which conventional search algorithms are not tied.

Reinforcement learning is the problem of agents to learn from the environment by their interactions with dynamic environment. We can relate them to the learning

agents. The interactions are trial and error in nature because a supervisor does not tell the agent which actions are right or wrong, unlike the case in supervised learning. There are mainly two main strategies to solve this problem. The first one is to search in behavioral space to find out the action behavior pair that performs well in the environment. The other strategy is based on statistical techniques and dynamic programming to estimate the utility of actions and chances of reaching a goal.

1.9 REINFORCEMENT FUNCTION AND ENVIRONMENT FUNCTION

As discussed above, the reinforcement learning is not just the exploitation of information based on already acquired knowledge. Rather, reinforcement learning is about balance between exploitation and exploration. Here exploitation refers to making the best use of knowledge acquired so far, while exploration refers to exploring new action, avenues, and route to build new knowledge. While exploring the action is performed, each action leads to learning through either rewards or penalties. The value function is the cumulative effect, while reward is associated with a particular atomic action. The environment needs to be modeled in changing scenarios so that it can provide the correct response that can optimize the value. The reinforcement function here is the effect of the environment, which allows the reinforcement.

Figure 1.6 depicts a typical reinforcement-learning scenario where the actions lead to rewards from environment. The purpose is to maximize expected discounted returns and also called value. The expected returns are given by

$$E\{r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots\}$$

Here discount rate is $0 \leq \gamma \leq 1$.

Finally the value of being in state s with reference to policy P_i is of interest to us and is calculated by

$$V^{\pi}(s) = E_{\pi}\{r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots / s_t = s\}$$

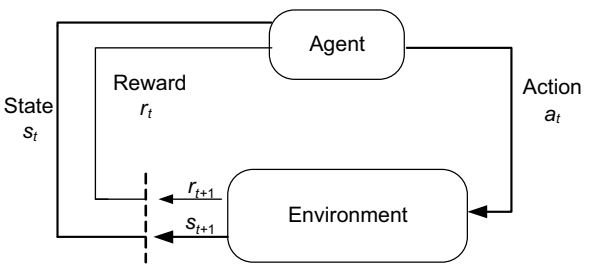

Figure 1.6 Reinforcement-learning scenario.

In short, for every action, there are environment functions and reinforcement functions. We will deal with these functions in greater detail as we proceed in this book.

1.10 NEED OF REINFORCEMENT LEARNING

Neither exploration nor exploitation alone can exhibit the intelligent learning behavior that is expected in many real-life and complex problems. A technique that makes use of both is required. While a child is learning to walk, it makes use of supervised as well as unsupervised ways of learning. Supervised here refers to inputs given to a child by parents while it may try to classify objects based on similarity and differences. Furthermore, a child explores new information by a new action and registers it. That even happens simultaneously. While kids are exploiting the knowledge, they also explore their outcomes with new actions, register learning, and build knowledge base, which is exploited in the future. In fact, exploration along with environment and learning based on the rewards or penalties is required to exhibit the intelligent behavior expected in most of the real-life scenarios.

Take an example of an intelligent automated boxing trainer. The trainer needs to exhibit more and more intelligence as it progresses and comes across a number of boxers. In addition, the trainer needs to adapt to a novice as well as expert. Furthermore, the trainer also needs to enhance his/her performance as the candidate starts exhibiting better performance. This very typical learning behavior is captured in reinforcement learning and hence it is necessary to solve many real-life problems. Learning based on data and perceived data pattern is very common. At any point of time, intelligent systems act based on percept or sequence of percepts. The percept here is the view of the intelligent system about the environment. Effective learning based on percept is required for real-time and dynamic intelligent systems. Hence machine intelligence needs learning with reference to environment, explores new paths, and exhibits intelligence in known or new scenarios. Reinforcement learning captures these requirements; hence for dynamic scenarios, reinforcement learning can be used effectively.

1.11 REINFORCEMENT LEARNING AND MACHINE INTELLIGENCE

The changing environment and environmental parameters and dynamic scenarios of many real-life problems make it difficult for a machine to arrive at solutions. If a computer could learn to solve the problems—through exploration or through trial and error—that would be of great practical value. Furthermore, there are many situations where we do not know enough about environment or problem scenarios to build an expert system, and even the correct answers are not known. The examples are car control, flight control, and so on, where there are many unknown parameters and scenarios. “Learning how to achieve the goal without knowing the exact goal till it achieves the goal” is the most complex problem intelligent systems are facing.

Reinforcement learning has one of the most important advantages for these types of problems, that is, advantage of updating.

Every moment there is a change in scenarios, and environmental parameters in the case of dynamic real-life problems. Take an example of a missile trying to hit a moving target, an automatic car driver, and business intelligent systems—in all these cases the most important aspect is learning from exploration and sensing the environment response for every progressive action. The information about the goal is revealed as we explore with help of new actions. This learning paradigm helps us in reaching a goal without the prior knowledge about the route or similar situations.

1.12 WHAT IS SYSTEMIC LEARNING?

As we have discussed above, in a dynamic scenario the role of environment and the interactions of the learning agent with environment become more and more important. Interestingly, it is an important thing to determine environment boundaries and understand the rewards and penalties of any action with reference to environment. As it becomes more and more complex and becomes more and more difficult, it also becomes very important to define the environment in dynamic scenarios. Furthermore, it even becomes necessary to understand the impact of any action from a holistic perspective. The sequence of percepts with reference to a system may need to be considered in this case. That makes it necessary to learn systemically. The fact is that sometimes rewards may not be immediate while it might be necessary to take into account the system interactions with reference to an action. The rewards, penalties, and even the resultant value need to be calculated systemically. To exhibit systemic decision making, there is a need to learn in a systemic way. The capture and building percept with all system inputs and within system boundaries is required.

Learning with a complete system in mind with reference to interactions among the systems and subsystems with proper understanding of systemic boundaries is systemic learning. Hence the dynamic behavior and possible interactions among the parts of a subsystem can define the real rewards for any action. This makes it necessary to learn in a systemic way.

1.13 WHAT IS SYSTEMIC MACHINE LEARNING?

Making machines to learn in a systemic way is systemic machine learning. Learning in isolation is incomplete—furthermore, there is no way to understand the impact of actions on environment and long-term prospects of reaching a goal. But the other aspect of systemic machine learning is to understand the system boundaries, determine the system interactions, and also try to visualize the impact of any action on the system and subsystems. The systemic knowledge building is more about building holistic knowledge. Hence it is not possible with an isolated agent, but rather it is the organization of intelligent agents sensing the environment in various ways to understand the impact of any action with reference to environment. That further

leads to building the holistic understanding and then deciding the best possible action based on systemic rewards received and inferred. The system boundaries keep changing and the environment function in traditional learning fails to explore in multiobjective complex scenarios. Furthermore, there is a need to create the systemic view, and systemic machine learning tries to build this systemic view and make the system learn so that it can be capable of systemic decision making. We will discuss various aspects of systemic learning in Chapters 2 and 3.

1.14 CHALLENGES IN SYSTEMIC MACHINE LEARNING

Learning systemically can solve many real-life problems available at hand. But it is not easy to make machines learn systemically. It is easy to develop learning systems that work in isolation, but for systemic learning systems there is a need to capture many views and knowledge about the system. For many intelligent systems just based on percept or rather a sequence of percepts, it is not possible to build a system view. Furthermore, to solve the problems and simplify the problems to represent a system view, there is a need to go ahead with a few assumptions, and some of these assumptions do not allow us to build the system view in the best possible way. To deal with many complexities in systemic machine learning, we need to go for complex models; and in the absence of knowledge about the goal, the decisions about the assumptions become very tricky.

In systemic thinking theory the cause and effects can be separated in time and space, and hence understanding impact of any action within the system is not an easy task. For example, in the case of some medicine we cannot see the results immediately. While understanding the impact of this action, we need to decide time and system boundaries. With any action the agent changes its state and so does the system and subsystem. Mapping these state transitions to the actions is one of the biggest challenges. Other challenges include limited information, understanding and determining system boundaries, capturing systemic information, and systemic knowledge building. In subsequent chapters the paradigm of systemic learning with the challenges and means to overcome them are discussed in greater detail.

1.15 REINFORCEMENT MACHINE LEARNING AND SYSTEMIC MACHINE LEARNING

There are some similarities between reinforcement learning and systemic machine learning while there are subtle differences. Interestingly, reinforcement learning and systemic machine learning is based on a similar foundation of exploration in a dynamic scenario. Furthermore, reinforcement learning is still more goal centric while systemic learning is holistic. The concept of systemic machine learning deals with exploration, but more thrust is on understanding a system and the impact of any action on the system. The reward and value calculation in systemic machine learning is much more complex. Systemic learning represents the reward from the system as

the system reward function. The reward it gets from various subsystems and the cumulative effect is represented as the reward for an action. Another important thing is inferred reward. Systemic machine learning is not only exploration, and hence the rewards are inferred. This inference is not limited to the current state, but it also inferred for n states from the current state. Here n is the period of inference. As the cause and effects can be separated in time and space, rewards are accumulated across the system and inferred rewards are accumulated from the future states.

1.16 CASE STUDY PROBLEM DETECTION IN A VEHICLE

As discussed in great detail in the next chapter, a system consists of interrelated parts that together work to create value. A car is a system. When there is startup trouble in a car, it is advised to change the ignition system. In reinforcement learning, you change the ignition and the car starts working fine; after 8–10 days the car again starts giving the same problem. It is taken to mechanic who changes an ignition system again. This time he uses an ignition system of better quality. The issue is resolved and you receive positive reward. Again after a week or so the car begins giving startup trouble once again. Taking into account the whole system can help to solve these types of problems. The central locking system that was installed before this problem occurred is actually causing the issue. The impact on the whole system due to the central locking system is not considered previously, and hence the problem remains unattended and unresolved. As we can see here, the cause and effects are separated in time and space and hence no one has looked at the central locking system. In systemic machine learning, considering the car as a system, the impact of central locking is checked with reference to a complete system, that is, complete car and hence the problem can be resolved in a better way.

1.17 SUMMARY

Decision making is a complex function. Day by day the expectations from the intelligent systems are increasing. Isolated and data-based intelligence can no longer meet expectations of the users. There is a need to solve the complex decision problems. To do this, there is a need to exploit the existing knowledge and also explore new routes and avenues. This happens in association with environment. For any action the environment provides the reward. The cumulative reward is used in reinforcement learning to decide actions. Reinforcement learning is like learning with critic. Once an action is performed, a critic criticizes it and provides feedback. Reinforcement learning is extremely useful in dynamic and changing scenarios such as boxing training, football training, and business intelligence.

Although reinforcement learning is very useful and captures the essence of many complex problems, the real-world problems are more systemic in nature. Furthermore, one of the basic principles of systemic behavior is that cause and effects are separated in time and space. It is very true for many real-life problems. There is a need

of systemic decision making to solve these complex problems. To make systemic decisions, there is a need to learn systemically. Systemic machine learning involves making a machine learn systemically. To learn systemically, there is need to understand system boundaries, interaction among subsystems and impact of any action with reference to a system. The system impact function is used to determine this impact. With broader and holistic system knowledge, it can deal with complex decision problems in a more organized way to provide best decisions.

REFERENCE

1. Senge P. *The Fifth Discipline—The Art & Practice of The Learning Organization*. Currency Doubleday, New York, 1990.