Chapter 1

Monday, December 17, 2018 10:40 AM

Page 1 to 5:

Introduction:

Humans learn by interacting with environment. When an infant plays, waves its arms or looks about, it has a direct sensorimotor connection to its environment and this connection produces information about the cause and effect, consequences of actions and also what to do in order to achieve goals.
In everyday life we are aware about how the environment reacts to what we do and we try to influence what happens through the behavior. Learning from interaction is a foundational idea in nearly all theories of learning and intelligence.
Rather than directly theorizing about how people or animals learn, the book explores idealized learning situations and evaluate the effectiveness of various learning methods.
The book explores designs for machines that are effective in solving learning problems, evaluating the designs via mathematical analysis or computational experiments. So, the book is more focused on goal-directed learning from interaction.

1.1. Reinforcement Learning

RL is learning what-to-do (how to map situations to actions) to maximize a numeric reward signal. The learner has no idea about which actions to take, but it has to discover which actions yield the most reward by trying them. But in some cases the actions not only affect the immediate reward but also the next situation and through that all the subsequent rewards.

Trial-and-error search and delayed reward are the two important characteristics and distinguishing features of RL.

The RL problem can be formalized using the ideas from dynamical systems theory , like optimal control for the incompletely known Markov Decision processes. Markov Decision processes are intended to include just three aspects - sensation, action and goal.

The learning agent should be able to sense the state of its environment to some extent and must be able to take actions that affect the state and also the agent should have a goal or goals relating to the state of the environment.

RL is different from Supervised Learning in the fact that the Supervised learning is provided with the labeled examples (labeled by a knowledgeable external supervisor) and models will train from them.

RL is also different from the Unsupervised Learning(USL) in the fact that the USL is typically about finding the structure hidden within the collections unlabeled data. But RL is trying to maximize a reward signal instead of trying to find the hidden structure.

Of all the forms of machine learning, reinforcement learning is the closest to the kind of learning that humans and other animals do, and many of the core algorithms of reinforcement learning were originally inspired by biological learning systems.

Uncovering the hidden structure in an agent's experience certainly be useful in reinforcement learning, but by itself it does not address the problem of maximizing a reward signal.

Hence the RL is considered as a new paradigm apart from Supervised and Unsupervised learning.

The challenges that arise in RL is the trade-off between exploration and exploitation.
Exploitation:
To get a lot of reward the agent has to prefer actions that it has tried before and found to be effective in producing reward. This is called as exploitation and the agent has to do it to obtain reward

Exploration:
But in order to discover the actions that offer good reward, the agent has to try actions that it has not selected before and find them . This is called as exploration and it is needed to make better action selections in the future.

The agent must try a variety of actions and progressively favor 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.

RL explicitly considers the whole problem of a goal-directed agent interacting with an uncertain environment. But much of the machine learning research concerned with supervised learning divide that big problem into a lot of sub-problems, the focus on isolated subproblems is the limitation for this approach.

RL agents have explicit goals, can sense the environments, and can choose actions to influence their environments.

1.2. Examples

Some of the examples that explain the reinforcement learning:

1. Chess player makes a move:
   The choice is informed by planning (anticipating possible replies and counter replies) and by immediate judgements of the desirability of particular positions and moves.
2. An adaptive controller adjusts the parameters of refinery's operation in real time:
   It optimizes the yield/ cost / quality trade-off on the basis of specified marginal costs without strictly following the set points originally set by engineers.
3. A gazelle calf learning to run after half an hour it is born
   The calf will struggle to its feet minutes after being born and it will learn using trial and error

All of the above examples involve the interaction between an active - decision making agent and its environment, within which the agent seeks to achieve a goal despite uncertainty about its environment. The agents actions affect the future state of the environment (e.g., the next chess position, the level of reservoirs in refinery), which will affect the actions and opportunities available to the agent at the later times. Foresight and planning, which take into account the indirect and delayed consequences of actions, are required to choose correct actions. The agents can judge its progress towards the goal based on what it can sense directly.
The knowledge the agent brings to the task at the start—either from previous experience with related tasks or built into it by design or evolution—influences what is useful or easy to learn, but interaction with the environment is essential for adjusting behavior to exploit specific features of the task.

1.3. Elements of Reinforcement Learning

Beyond the agent and the environment there are four main sub-elements of a reinforcement learning system:

1. A policy
2. A reward signal
3. A value function
4. A model of the environment (optional)

Policy:

1. Agent's way of behaving at a given time.
2. Mapping from perceived states of the environment to actions to be taken
3. In terms of Psychology set of stimulus-response rules
4. May be a simple function or lookup table or extensive computation such as search process
5. It is the core of a reinforcement learning agent it alone can be sufficient to determine behavior
6. Generally, policies may be stochastic, specifying probabilities for each action

Reward Signal:

1. Defines the goal of a reinforcement learning problem
2. On each time step the environment provides the agent a single number called reward
3. Agent's sole objective is to maximize the total reward it receives over the long run
4. It denotes the events if it is good or bad for the agent
5. Reward is analogous to the pain or pleasure in biological systems
6. It is the primary basis for altering the policy
7. Reward signals are the stochastic functions of the state of the environment and the actions taken

Value Function:

1. Specifies what is good in long run
2. Value of a state is the total amount of reward an agent can expect to accumulate over the future starting from that state
3. Values indicate the long-term desirability of states taking into account the states that are likely to follow and rewards available in those states.
4. Rewards are in a sense primary, whereas values, as predictions of rewards, are secondary.
5. Only purpose of estimating values is to achieve more reward
6. Value are the most important ones in making and evaluating the decisions
7. Action choices are made based on value judgements, that bring states of highest value because these actions obtain the greatest amount of reward over the long run
8. Values are hard to be determined because it has to be estimated and re-estimated from the sequences of observations an agent makes over its entire lifetime
9. Method for efficiently estimating the values is an important one

Model of the environment (Optional):

1. It mimics the behavior of the environment
2. Given a state and an action the model might predict the resultant next state and next reward
3. Models are used for planning, they provide the way of deciding on a course of action by considering possible future situations before they are actually experienced.
4. Methods for solving reinforcement problems that use models and planning are called model-based methods, opposed to simpler model-free methods that ae explicitly trial- and-error learners.

1.4. Limitations and Scope

Reinforcement learning relies heavily on the states. They are passed as input to the policy and value function and as both input to and output from the model. State can be considered as a signal conveying to the agent some sense of "how the environment is" at a particular time.
Most of the reinforcement learning methods considered in this book are structured around estimating the value functions, but it is not necessarily to do the same way. Other methods such as genetic algorithms, genetic programming, simulated annealing and other optimization methods never estimate value functions.
In algorithms like the one above, multiple static policies interacting over an extended period of time with a separate instance of the environment. The policies that perform well and getting the most reward and random variations of them are carried over to the next generation of policies and the

process is repeated. These are called as evolutionary methods because they are analogous to the way biological evolution produces organisms with skilled behavior. These methods are particularly effective if the space of policies is sufficiently small or can be structures so that the good policies are common or easy to find. They are also having advantage on problems in which the learning agent cannot sense the complete state of its environment.

Evolutionary methods won't learn while interacting with the environment. But the ones that learn when interacting with the environment are much more efficient than the evolutionary methods. Because evolutionary methods ignore much of the useful structure of the environment. They do not use the fact that the policy they are searching for is a functions from states to actions and also they do not notice which states an individual passes through during its lifetime or which action it selects.

But evolutionary methods along with the learning methods work together greatly.