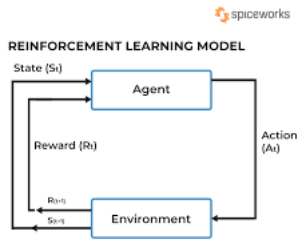
Reinforcement learning

This is somewhere between supervised and unsupervised learning. The algorithm gets told when the answer is wrong, but does not get told how to correct it. It has to explore and try out different possibilities until it works out how to get the answer right. Reinforcement learning is sometime called learning with a critic because of this monitor that scores the answer, but does not suggest improvements. Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards. A learner (the program) is not told what actions to take as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situations and, through that, all subsequent rewards.

Example:

Consider teaching a dog a new trick: we cannot tell it what to do, but we can reward/punish it if it does the right/wrong thing. It has to find out what it did that made it get the reward/punishment. We can use a similar method to train computers to do many tasks, such as playing backgammon or chess, scheduling jobs, and controlling robot limbs. Reinforcement learning is different from supervised learning. Supervised learning is learning from examples provided by a knowledgeable expert.



Reinforcement learning This is somewhere between supervised and unsupervised learning. The algorithm gets told when the answer is wrong, but does not get told how to correct it. It has to explore and try out different possibilities until it works out how to get the answer right. Reinforcement learning is sometime called learning with a critic because of this monitor that scores the answer, but does not suggest improvements. Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards. A learner (the program) is not told what actions to take as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them. In the most interesting and challenging cases, actions may affect not only the immediate reward but also the next situations and, through that, all subsequent rewards. Example Consider teaching a dog a new trick: we cannot tell it what to do, but we can reward/punish it if it does the right/wrong thing. It has to find out what it did that made it get the reward/punishment. We can use a similar method to train computers to do many tasks, such as playing backgammon or chess, scheduling jobs, and controlling robot limbs. Reinforcement learning is different from supervised learning. Supervised learning is learning from examples provided by a knowledgeable expert.

1. Uncertainty plays a central role in reinforcement learning. The agent's environment

and its own behavior can be subject to random fluctuations so that the outcomes

of decisions cannot be known beforehand with complete certainty. An accurate probabilistic model of the these uncertainties may, or may not, be available to the agent.

2. The reward input to the agent can be any scalar signal evaluating the agent's behavior. It might indicate just success when a goal state is reached, just failure while not reaching a goal state, or it might provide moment-by-moment evaluatuations of on-going behavior (as, for example, in giving the amount of energy currently being consumed while a task is being accomplished). Moreover, multiple evaluation criteria can be combined in various ways to form the scalar reward signal (for example, via a weighted sum).

3. An important difficulty faced by a reinforcement learning system is the creditassignment problem (Minsky 1961): How do you distribute credit for success among the many decisions that may have been involved in producing it? (See also REINFORCEMENT LEARNING IN MOTOR CONTROL.)

4. A reinforcement learning system often has to forgo immediate reward in order to obtain more reward later or over the long run. This kind of "sacrificing" behavior arises because the agent's actions influence not only each reward input but also the environment's state transitions. An action may be preferred because it sets the stage for a large reward later rather for its immediate reward.

5. The reward signal does not directly tell the agent what action is best; it only evaluates the action taken. A reward input also does not directly tell the agent how to change its actions. These are key features distinguishing reinforcement learning from supervised learning, and we discuss them further below.

6. Reinforcement learning algorithms are selectional processes. There must be variety in the action-generation process so that the consequences of alternative actions can

be compared to select the best. Behavioral variety is called exploration; it is often generated through randomness, but it need not be.

7. Reinforcement learning involves a conflict between exploitation and exploration. In deciding which action to take, the agent has to balance two conflicting objectives: it has to exploit what it has already learned to obtain high rewards, and it has to behave in new ways—explore—to learn more. Because these needs ordinarily conflict, reinforcement learning systems have to somehow balance them. In control engineering, this is known as the conflict between control and identification.

8. Some researchers think of reinforcement learning as a form of supervised learning (because the reward input is a kind of supervision), and others think of it as a form of unsupervised learning (because the reward input is not like the label of an example). There is some truth to each of these views, but reinforcement learning is really different from both. A key distinguishing feature is the presence in reinforcement learning of the conflict between exploitation and exploration. This is absent from supervised and unsupervised learning unless the learning system is also engaged in influencing which training examples it sees