Reinforcement learning

A sequential decision process (SDP) is any dynamic process that involves a time ordered sequence of decisions that influence some cost or reward that accrues over time. The idea is to determine the decision sequence that optimizes the defined cost or reward.

A third branch of machine learning, referred to as reinforcement learning, shares some of the essential characteristics of supervised and unsupervised learning, but is applied specifically to the problem of sequential decision making and therefore relies on a different form of data to that used by either supervised or unsupervised learning. Much of the theory of reinforcement learning is based on **dynamic programming**, which was formulated by Richard Bellman in the 1950s (Bellman, 1957). While this provides much of the mathematical foundation for reinforcement learning, as we will see below, dynamic programming has a limited computational capacity. Practically, this means that dynamic programming, at least in its exact form, is not a viable solution for applications made to exploit the type of high throughput data that is often available today.

One especially compelling early application of reinforcement learning was the development of an SDP, known as TD-Gammon, which is able to play backgammon at a competitive level. More recently, reinforcement learning was used by DeepMind Technologies to create a chess-playing program known as AlphaZero. Reinforcement learning is particularly suited to develop game-playing applications, especially when coupled with deep neural network models (deep reinforcement learning). For example, it is currently used to create quite effective video game playing applications.