**Reinforcement Learning**

Ein sequentieller Entscheidungsprozess (Sequential Decision Process, SDP) ist ein dynamischer Prozess, in dem eine zeitlich geordnete Abfolge von Entscheidungen Auswirkungen auf bestimmte Kosten oder Belohnungen hat, die sich im Laufe der Zeit verstärken. Es geht darum, die Entscheidungssequenz zu bestimmen, die die Kosten oder die Belohnung, um die es geht, optimiert.

Reinforcement Learning ist der dritte der drei Zweige von maschinellem Lernen. Es weist einige wichtige Gemeinsamkeiten mit Supervised und Unsupervised Learning auf, wird jedoch speziell für die Problemstellung sequentieller Entscheidungsfindung eingesetzt und beruht daher auf einer anderen Form von Daten als denen, auf die Supervised und Unsupervised Learning zurückgreifen. Ein Großteil der Theorie zum Reinforcement Learning beruht auf der **dynamischen Programmierung**, einem von Richard Bellman in den 50er Jahren geprägten Konzept (Bellman, 1957). Obwohl die dynamische Programmierung, wie unten gezeigt wird, eine wesentliche mathematische Grundlage des Reinforcement Learning bildet, bietet sie nur begrenzte Rechenkapazität. In der Praxis bedeutet das, dass dynamische Programmierung, zumindest in ihrer strengen Form, keine tragfähige Lösung für Anwendungen darstellt, die die heute oft vorhandenen Hochdurchsatzdaten auswerten sollen.

Eine besonders überzeugende Anwendung von Reinforcement Learning war die Entwicklung eines als TD-Gammon bekannten SDP, der fähig ist, Backgammon auf Wettkampfniveau zu spielen. In neuerer Zeit wurde Reinforcement Learning von DeepMind Technologies für die Entwicklung des Schachspielprogramms AlphaZero eingesetzt. Reinforcement Learning eignet sich besonders gut für die Entwicklung von Spiele-Anwendungen, insbesondere in Verbindung mit tiefen neuronalen Netzmodellen (Deep Reinforcement Learning). Gegenwärtig wird es beispielsweise eingesetzt, um durchaus effiziente Videospielanwendungen zu entwickeln.