

## Bestärkendes Lernen

Ein sequenzieller Entscheidungsprozess (SEP) ist jeder dynamische Prozess, der eine zeitgeordnete Sequenz von Entscheidungen beinhaltet, die manche der mit der Zeit anfallenden Verluste oder Gewinne beeinflussen. Die Idee dahinter ist es, die Entscheidungssequenz zu ermitteln, welche die festgelegten Verluste oder Gewinne optimiert.

Ein drittes Teilgebiet des Maschinenlernens, das bestärkende Lernen, teilt manche der grundlegenden Merkmale des überwachten und unüberwachten Lernens. Es wird jedoch speziell auf das Problem der sequenziellen Entscheidungsfindung angewandt und baut daher auf einer anderen Art von Datensätzen auf, als es bei überwachtem und unüberwachtem Lernen der Fall ist.

Ein Großteil der Theorie des bestärkenden Lernens basiert auf dem Konzept der 1950 von Richard Bellman formulierten dynamischen Programmierung (Bellman, 1957). Während diese die mehrheitliche mathematische Grundlage für bestärkendes Lernen ausmacht, werden wir später auf die begrenzte Rechenkapazität der dynamischen Programmierung zurückkommen. In der Praxis bedeutet sie die Unzulänglichkeit der klassischen dynamischen Programmierung für Anwendungen, die darauf abzielen heutzutage verfügbare, hohe Datendurchläufe zu nutzen.

Eine frühe relevante Anwendung des bestärkenden Lernens, war die Entwicklung eines SEPs namens TD-Gammon, welcher dazu fähig ist, Backgammon auf Wettkampfniveau zu spielen. Ein gegenwertiges Beispiel ist AlphaZero, ein Schachspielprogramm, das von DeepMind Technologies durch die Nutzung bestärkenden Lernens entwickelt wurde.

Bestärkendes Lernen ist vor allem dann geeignet, Gameplay Applikationen zu entwickeln, wenn es mit tiefen neuronalen Ansätzen (tiefbestärkendes Lernen) verknüpft wird. Beispielsweise wird es derzeit genutzt, um effektive Video-Gameplay-Applikationen zu entwickeln.