Hochschule Bremerhaven

Exposé für eine Bachelorarbeit zum Thema:

Reinforcement Learning

Theoretische Grundlagen und praktische Umsetzung am Beispiel eines Ameisen-Agentenspiels oder eines lernenden Systems zur Bestimmung des optimalen Lenkverhaltens für einen autonom agierenden Einparkassistenten

Autor:	Jan Löwenstrom
Matrikelnr.:	34937
Erstprüfer:	Prof. Dr.-Ing. Henrik Lipskoch
Zweitprüfer:	Prof. Dr. Mathias Lindemann || Prof. Dr. Nadija Syrjakow

6. Januar 2020

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Das ist eine Einleitung
(Fedjaev, 2017)

2 Markow Entscheidungsprozess

Ein Markov Descision Process

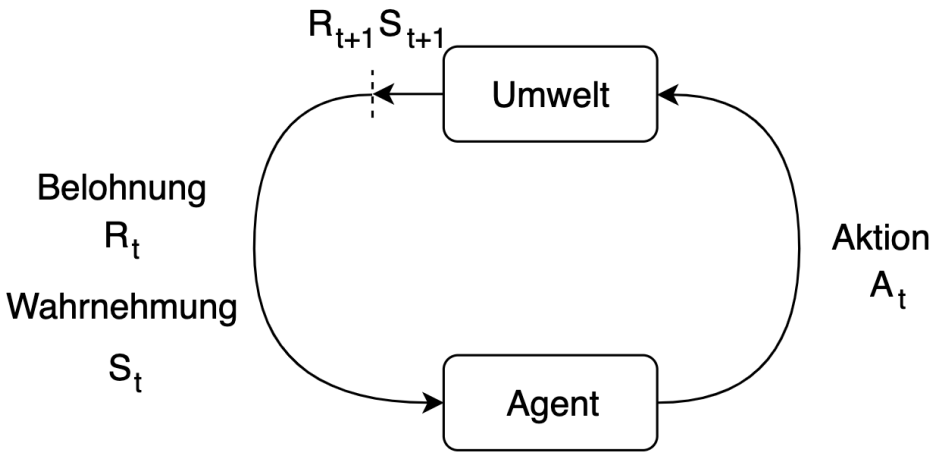

Abbildung 1: Agent-Umwelt Interface

Der Agent interagiert mit dem MDP jeweils zu diskreten Zeitpunkten $t = 0, 1, 2, 3, ...$ Zu jedem Zeitpunkt t beobachtet der Agent den Zustand seiner Umgebung $S_t \in \mathcal{S}$ und wählt aufgrund dessen eine Aktionen $A_t \in \mathcal{A}$. Als Konsequenz seiner Aktion erhält er einen Zeitpunkt später eine Belohnung $R_{t+1} \in \mathcal{E} \subset \mathbb{R}$ und stellt den Folgezustand S_{t+1} fest. Das Zusammenspiel zwischen Agenten und MDP erzeugt also folgende Reihenfolge:

$$S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3, ...$$

3 Markow-Eigenschaft und Zustandsmodellierung

Die Markow-Eigenschaft, obwohl relativ simpel, erhält ein eigenes Kapitel, da sie von fundamentaler Wichtigkeit ist und bei der Modellierung eines Reinforcement Learning Problems eine besondere Rolle spielt. Verbinden lässt sich dies sehr gut mit einem Einblick über die generelle Modellierung von Zuständen bei einem RL Problem.

> The future is independent of the past given the present

Dieser Satz erscheint oft in Büchern und Papern, wenn es um die Markow-Eigenschaft geht, denn er versucht zusammenzufassen, was diese aussagt. Im Zusammenhang von MDPs lässt sich dieser Satz so übersetzen, dass ein Folgezustand nicht abhängig von Aktionen bzw. Zuständen in der Vergangenheit ist, sondern ausschließlich von dem aktuellen Zustand und der aktuell gewählten Aktion. In der Literatur gibt es unterschiedliche Auffassungen darüber, ob die Markow-Eigenschaft an den MDP direkt geknüpft ist oder an den Zustand, den der Agent zur Abwägung der Entscheidung zur Verfügung hat. Bei der ersten Annahme wird davon ausgegangen, dass der Zustand, der von der Umwelt ausgeliefert wird direkt die Markow-Eigenschaft besitzen muss. (Sutton S.49) hingegen bindet die Eigenschaft an den Zustand und nicht an den Entscheidungsprozess als solches. Ein Zustand ist somit die Menge aller notwendigen Informationen der Vergangenheit, die für die Zukunft relevant sind. Statt den gegebenen Zustand der Umwelt direkt zu übernehmen, werden hier Beobachtungen der Umwelt zu einer internen Repräsentation von Markow-Zuständen verarbeitet.

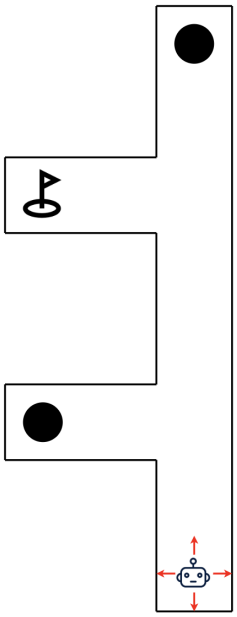

Abbildung 2: Zwei-Wege Beispiel zu der Markow-Eigenschaft

Literatur

Fedjaev, J. (2017). *Decoding eeg brain signals using recurrent neural networks.* Zugriff auf https://www.spektrum.de/kolumne/eine-waffe-gegen-malaria/1525481

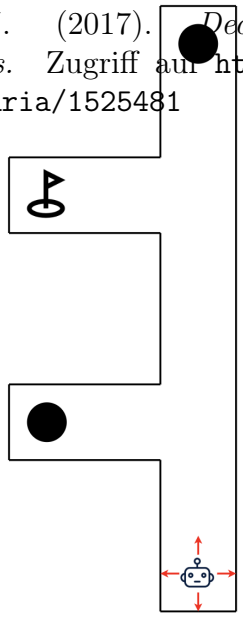

Abbildung 3: Birds