ZENTRUM FÜR MASCHINELLES LERNEN

HOCHSCHULE HEILBRONN

Airhockey auf übermenschlichen Level mit Deep Reinforcement Learning

Rico Steinke, Pascal Graf und Nicolaj C. Stache
Automotive Systems Engineering, Heilbronn University of Applied Sciences

Projektübersicht

Motivation:

- Der Demonstrator zeigt, wie eine künstliche Intelligenz für Anwendungen im Bereich der Robotik ohne menschliches Expertenwissen zunächst in einer Simulation trainiert und dann in der Realität nutzbar gemacht werden kann.
- Für komplexe Szenarien, wie das Spielen am Airhockey-Tisch, ist es äußerst schwierig, eine Handlungsheuristik manuell zu programmieren. Deshalb trainiert ein sogenannter Software-Agent beim Reinforcement Learning im Spiel gegen sich selbst, seine Spielstrategie zu optimieren.

Ansatz / Ziele:

- Eine künstliche Intelligenz lernt im Duell gegen sich selbst, das Geschicklichkeits- und Reaktionsspiel Airhockey zu spielen. Ziel jenes Spiels ist es, den luftgelagerten Puck mithilfe des eigenen Pushers in das gegnerische Tor zu schießen.
- Dabei erhält der Agent zu jedem Zeitschritt eine numerische Belohnung, welche über hunderttausende Spieleepisoden maximiert wird.

Übertragbarkeit auf reale Aufgaben:

- Das gleiche Prinzip lässt sich auch auf Probleme im Bereich der industriellen Produktion oder des autonomen Fahrens anwenden.
- Beispiele hierfür sind:
  - Anlagenoptimierung zur effizienteren Nutzung von Zeit, Ressourcen und Energie abhängig von Umwelteinflüssen
  - Feinmotorisches Greifen und Handling komplexer Objekte in der Robotik (Griff in die Kiste).
  - Automatisches Bremsen und Beschleunigen eines Fahrzeugs in Reaktion auf umliegenden Verkehr und unvorhergesehene Ereignisse

Airhockey-Simulation

- Die Simulation basiert auf der Spieleengine *Unity3D* [1] in Kombination mit der hochrealistischen Physik-Library *MuJoCo* (Multi-Joint dynamics with Contact) [2].
- *MuJoCo* ermöglicht eine genaue Parametrisierung Reibungs- und Kontaktverhaltens, was eine bessere Übertragbarkeit auf die Realität gewährleistet.
- Mit entsprechender Rechenkapazität können mehrere Spielinstanzen gleichzeitig simuliert und zeitlich beschleunigt werden, sodass pro Spielepisode im Schnitt weniger als zwei Sekunden Realzeit vergeht.

Abbildung 1: 3D-Airhockey-Simulation aus Spielerperspektive. Das User Interface zeigt Informationen zum Reinforcement Learning, der Steuerung sowie der Zusammensetzung der Belohnung, welche der Agent erhalten hat.

Reales Modell

- Der reale Demonstrator ist mit einer Kinematik ausgestattet, die über Schrittmotoren eine hochpräzise sowie dynamische Ansteuerung des Pushers erlaubt.
- Über dem Spielfeld ist eine High-Speed-Kamera angebracht, die das Spielfeld 150 mal pro Sekunde aufnimmt.
- Aus den Kamerabildern werden sowohl Positionen als auch Geschwindigkeiten von den relevanten Spielelementen (Pusher und Puck) extrahiert.
- Diese Informationen benötigt die KI, um sinnvolle Entscheidungen treffen zu können.

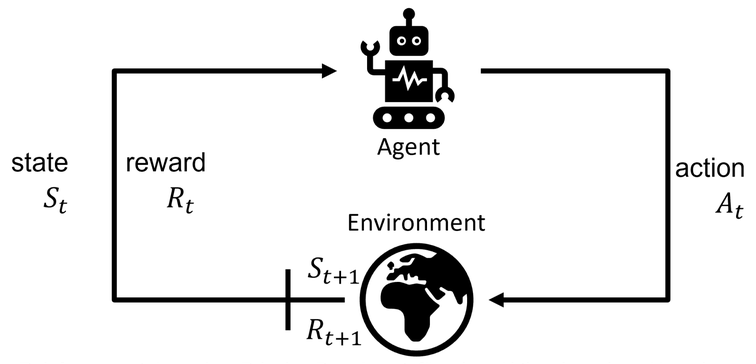

Abbildung : Reinforcement Learning Prinzip: Der Agent für eine Aktion in seiner Umgebung aus und erhält als Rückmeldung einen neuen State zusammen mit einem numerischen Reward.

Reinforcement Learning

Soft-Actor-Critic-Algorithmus [3]:

- Dieser Reinforcement Learning Algorithmus basiert auf dem Einsatz von zwei tiefen neuronalen Netzwerken.
- Das erste Netzwerk, *Actor* genannt, bestimmt dabei die nächste Aktion des Agents, also die Bewegungsrichtung- und Geschwindigkeit des Pushers.
- Das zweite Netzwerk, welches als *Critic* bezeichnet wird, schätzt die Qualität der gewählten Aktionen anhand des bisher erhaltenen Rewards ab und verbessert so implizit den *Actor.*
- Während des Trainings muss die KI abwägen, ob sie nach der Spielstrategie handelt, die sie im Moment für am besten hält, oder ob sie etwas Neues ausprobiert, um dadurch womöglich bessere Strategien zu erforschen (Exploration-Exploitation-Dilemma)

Reward-Zusammensetzung:

- Neben der Belohnung bzw. Bestrafung für (Gegen-)Tore, sollen andere unerwünschte Verhaltensweisen mit in die Rewardstruktur einfließen.
- Dazu Gehören das Fahren in Banden, schnelle Änderungen der Bewegungsrichtung oder das Dulden des Pucks in der eigenen Spielhälfte

Trainingsergebnisse

- Welche Belohnung die KI im Spiel erhalten kann, hängt maßgeblich von der Stärke ihres Gegners ab.
- Um eine unabhängige Bewertung zu erhalten, wurde ein Rating-System ähnlich dem ELO-Rating anderer Spiele (z.B. Schach oder Online Multiplayer Games) entworfen.
- Dabei spielen unterschiedliche Agents (abhängig von Trainingsdauer und -konfiguration) in einem Turnier gegeneinander.
- Nach etwa 48h Training hat der Agent das durchschnittliche menschliche Spielerniveau erreicht. Der beste, bisher ungeschlagene Agent hat ein Training von 138h (knapp sechs Tagen) absolviert, was etwa 220.000 Spielen entspricht.

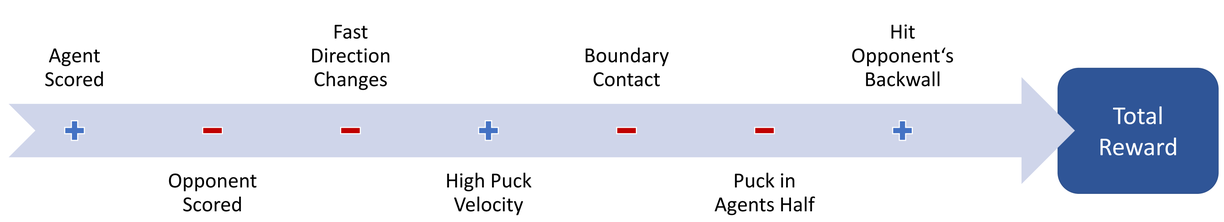

Abbildung 3: Zusammensetzung des Rewards, den die KI innerhalb einer Spieleepisode erhalten kann. Ziel des Reinforcement Learning Algorithmus‘ ist es, diesen zu maximieren.

Bewegungsanalyse

- Eine detaillierte Analyse der trainierten KI zeigt, dass diese das Abprallverhalten des Pucks über die Bande prognostizieren und sich dementsprechend frühzeitig positionieren kann.
- Über Heatmaps kann visualisiert werden, an welchen Positionen der Agent sich besonders häufig aufhält.

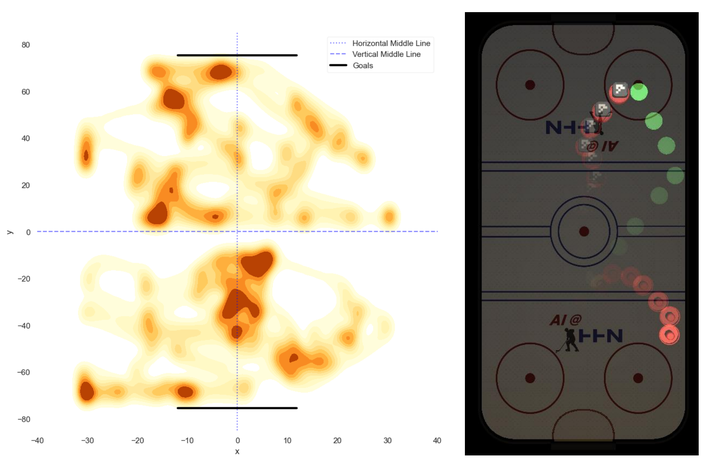

Abbildung 4: Mithilfe der Heatmaps (links) wird visualisiert, an welchen Positionen der Agent sich vorzugsweise aufhält. Die Bewegungstrajektorie (rechts) zeigt, wie der Agent gelernt hat, das Abprallverhalten des Pucks zu prognostizieren und diesen rechtzeitig abzufangen.

Quellen

[1] J. Haas, „A history of the unity game Engine,“ Worcester Polytechnic Institute, 2014.

[2] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 5026–5033, IEEE, 2012.

[3] T. Haarnoja et al., “Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor,” CoRR, 2018.

rsteinke@stud.hs-heilbronn.de, Pascal.Graf@hs-heilbronn.de, Nicolaj.Stache@hs-heilbronn.de