

Using Reinforcement Learning to Play Bomberman

Deep Reinforcement Learning (DRL) ist ein Bereich der künstlichen Intelligenz (KI), an dem sehr aktiv geforscht wird. DRL hat zum Ziel, einem Agenten beizubringen, sich in einer Umgebung zurechtzufinden. Deep Learning ist ein Teilbereich der KI, der in den letzten Jahren viele Probleme besser lösen konnte als klassische Methoden. Reinforcement Learning (RL) ist ein allgemeiner Ansatz für maschinelles Lernen mit dem Ziel, eine Steuerung zu lernen. Die Kombination der beiden Methoden ist DRL. Dieser wird das Potenzial zugeschrieben, generellere KI zu erschaffen.

In RL wird der aktuelle Zustand der Umgebung vom lernenden Agenten beobachtet. Führt der Agent eine Handlung aus, so kommt die Umgebung in einen neuen Zustand. Der Agent erhält ein Feedback für seine Handlung, welches als Reward (Belohnung) bezeichnet wird. Mit einem RL-Algorithmus lernt ein Agent, für jeden Zustand diejenige Handlung auszuführen, welche ihm den maximalen Reward erbringt. Dieser wiederholte Vorgang, etwas zu sehen, eine Handlung auszuführen, Feedback zu erhalten, eine neue Beobachtung zu erhalten und daraus mittels eines neuronalen Netzes Verhaltensverbesserungen abzuleiten, ist der Ansatz, welcher allen DRL-Methoden zugrunde liegt.

Das Ziel dieser Arbeit ist es, einem Agenten beizubringen, Pommerman, ein Klon des Spiels Bomberman, der für die Forschung mit DRL erstellt wurde, zu spielen. Pommerman wird von vier Agenten gleichzeitig gespielt. Es gibt verschiedene Modi. In Modus 'Free for All' gewinnt der Agent, der als längster überlebt; im Teammodus, bei welchem die vier Agenten in zwei Teams aufgeteilt sind, gewinnt das Team mit den letzten Überlebenden. Diese Arbeit untersucht zwei Ansätze. Als Erstes wurde der Multi-Agent Deep Deterministic Gradient (MADDPG) Algorithmus implementiert, welcher verspricht, Multi-Agent-Probleme gut lösen zu können. Dieser Versuch führt nicht zum Erfolg, was vermutlich daran liegt, dass MADDPG für Umgebungen mit kontinuierlichen Handlungen entwickelt wurde. In Pommerman können jedoch ausschliesslich diskrete Handlungen ausgeführt werden. Zweitens werden auch Ansätze, basierend auf Deep Q-Networks (DQN) getestet. Die Resultate zeigen, dass der double-DQN Algorithmus mit Verwendung eines Convolutional Neural Networks (CNN) in der Pommerman Umgebung Potenzial hat, wenn nur ein Agent lernt und die anderen eine fest programmierte Strategie verwenden.

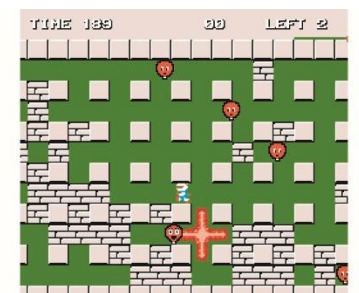


Diplomierende
Joram Liebeskind
Silvan Wehner

Dozierende
Thilo Stadelmann
Oliver Dürr



Screenshot von Pommerman, die Multi-Agent Reinforcement Learning Umgebung, welche in dieser Arbeit verwendet wurde.



Screenshot von Bomberman, einem arcade-ähnlichen Spiel, welches Pommerman für Multi-Agent Reinforcement Learning nachgebaut hat.
(Quelle: Wikipedia)