

Kurzzusammenfassung

In den letzten zehn Jahren hatte die Entwicklung von sogenannten KI Agenten bemerkenswerte Erfolge zu vermelden und erregte das Interesse der Öffentlichkeit, insbesondere im Bereich der strategischen Brettspiele wie Schach und Go. Diese Erfolge sind auf die intensive Forschung der Künstlichen Intelligenz in den verschiedensten Gebieten zurück zu führen, wo neuronale Netze ohne Zweifel einen großen Beitrag leisten. Aber auch die ständige Weiterentwicklung von Algorithmen im Bereich des *reinforcement learning* (RL) führte letztendlich zu einer neuen Methode, die unter dem Begriff *Monte-Carlo tree search* (MCTS) bekannt ist. Auf den ersten Blick sehr simpel, ist MCTS erstaunlich effektiv, insbesondere bei Problemen mit einem großen Zustandsraum. Obwohl strategische Brettspiele im Fokus von MCTS liegen, wurde es auch auf kombinatorische Optimierungsprobleme hin untersucht. Diese zeigten beeindruckende Ergebnisse, jedoch beruhen diese auf Methoden, die speziell auf solche Probleme zugeschnitten und somit nicht auf andere Probleme übertragbar sind.

Diese Arbeit untersucht im Kern die Lernparadigmen von KI Agenten auf kombinatorische Optimierungsprobleme, welche auf MCTS basieren. Anhand von RL Methoden und Monte-Carlo Algorithmen soll der KI Agent auf einen allgemeineren Fall erweitert werden, wodurch unser KI Agent auch auf ein weites Spektrum von Problemen anwendbar ist. Dieses Konzept nennt sich *general game playing* und stellt ein relativ neues und junges Forschungsfeld. Es ist auch mit der sogenannten *Black-Box Optimierung* verwandt. In diesem Zusammenhang stellen wir die Algorithmen UCT-ODU und dessen Erweiterung UCT-RAVE-ODU vor, beides KI Agenten, die dieses Konzept von einem universellen KI Agenten auf kombinatorische Optimierungsprobleme umsetzen sollen.

Die vorliegende Studie identifiziert die Schwierigkeiten, die entstehen wenn ein MCTS basierender KI Agent auf Optimierungsprobleme angewendet werden soll. Im Zusammenhang dazu definieren wir die Bedingungen, welche die Probleme erfüllen müssen und verdeutlichen die Notwendigkeit bestimmter Lernparadigmen. Schliesslich wird die Validierung unserer Vorgehensweise durch eine intensive empirische Ausarbeitung unterstützt und zeigen, dass die Algorithmen UCT-ODU und UCT-RAVE-ODU eine signifikante Verbesserung der Lernrate darstellen.