Zusammenfassung

In den vergangenen Jahren wurden Reinforcement Learning Agenten vermehrt zur Steigerung der Effizienz und Produktivität in Produktionssystemen eingesetzt. Bis dahin wurden regelbasierte Heuristiken verwendet, die jedoch bei komplexeren Problemen schnell an ihre Grenzen geraten und dadurch keine optimale Lösung finden können. Im Gegensatz dazu erzielen Reinforcement Learning Agenten durch ihre zustandsabhängige Flexibilität bessere Ergebnisse, insbesondere für komplexere Probleme. Allerdings beschränkte sich der Einsatz von Reinforcement Learning Techniken bisher größtenteils auf endliche Aufgaben. In der realen Welt steht man jedoch häufiger vor Problemen ohne klar definiertes Ende (kontinuierliche Probleme). Reward Machines beschreiben ein Konzept, das die Struktur der Reward-Funktion offenlegt und dieses Wissen nutzt, um den Trainingsprozess eines Reinforcement Learning Agentens zu beschleunigen. Dieses Konzept fokussiert sich jedoch ebenfalls hauptsächlich auf die Verwendung für endliche Aufgaben. Diese Masterarbeit untersucht, ob sich das Konzept der Reward Machines auch für kontinuierliche Aufgaben anwenden bzw. übertragen lässt. Einerseits beschäftigt sich diese Arbeit damit, ob das theoretische Konzept der Reward Machines eine Nutzung für kontinuierliche Aufgaben zulässt. Andererseits wird analysiert, ob der Algorithmus zur Beschleunigung des Trainings auch für unendliche Probleme funktioniert. Die Analyse des theoretischen Konzepts zeigt, dass Anpassungen notwendig sind, da Reward Machines als endliche Automaten nur endliche Eingabesequenzen erlauben. Deshalb sind Anpassungen insbesondere in der Definition der akzeptierenden Zustände notwendig. Diese Änderungen sind zwar notwendig, um eine korrekte Definition zu erhalten, jedoch haben sie nur wenig Einfluss auf das tatsächliche Training. Darüber hinaus werden Anpassungen am Algorithmus vorgenommen. Diese sind zwar technisch nicht zwingend notwendig, jedoch erscheinen sie durch den unendlichen Charakter kontinuierlicher Probleme sinnvoll. Diese Anpassungen werden in Form von Experimenten an Portalrobotersteuerungen, welche als kontinuierliches Problem einzustufen sind, getestet. Die Experimente zeigen, dass die Frage nach der Schnelligkeit des Lernens für die untersuchten Ansätze nicht leicht zu beantworten ist, da die Antwort maßgeblich durch die Definition von Geschwindigkeit beeinflusst ist. Einerseits verdeutlichen die Experimente, dass sowohl der ursprüngliche als auch der angepasste Algorithmus schneller eine Policy erlernen kann als ein Lernansatz ohne Reward Machines, wenn man die Schnelligkeit über die Anzahl der beobachteten Aktionen beziehungsweise die Anzahl der verwendeten Episoden zugrunde legt. Andererseits identifizieren die Experimente den Lernansatz ohne Reward Machine als effizienter in Bezug auf die Datennutzung, was auch als Aspekt der Geschwindigkeit betrachtet werden kann. Jedoch wird auch deutlich, dass der

ursprüngliche Ansatz Schwierigkeiten hat, den Umgang mit selten auftretenden Zuständen zu erlernen. Diese Zustände können durch stochastische Aspekte, wie z.B. Maschinenausfälle, beeinflusst sein. Im Gegensatz dazu ist der angepasste Algorithmus in der Lage, mit solchen Herausforderungen umzugehen.

Schlagworte - Reinforcement Learning, Reward Machine, Automatentheorie, Portalrobotersteuerung, Produktionssysteme

Abstract

During the last decades, reinforcement learning agents have found their way into production systems. They have been used to increase the efficiency and productivity of a factory, as reinforcement learning agents are able to outperform rule-based heuristics due to their great state-dependent flexibility. So far, most efforts have been made to apply reinforcement learning techniques to episodic tasks. However, most real-world problems are continuous problems that need to be solved. Reward machines represent a concept that reveals the structure of the reward function and exploits this knowledge to accelerate the training process of a reinforcement learning agent. Since this concept focuses on the use in episodic tasks, this master thesis investigates whether the concept of reward machines is applicable or transferable to continuous tasks. On the one hand, the research is concerned with whether the theoretical foundations of reward machines allows the use in continuous tasks. On the other hand, it is analysed whether the algorithm for exploitating the reward machines also works for continuous tasks. The analysis of the theoretical foundations shows that adjustments are necessary because reward machines, as finite state machines, only work for finite input sequences. Therefore, adjustments are required above all in the definition of acceptance states. However, these changes are necessary but only relevant for a sound definition and have little impact on the actual training. Furthermore, changes in the algorithm are not technically necessary, but seem reasonable due to the infinite nature of a continuous task. The experiments show that the question of speed of learning for the examined approaches cannot be easily answered, since the definition of speed is relevant. The original approach as well as proposed adaptions are tested by means of experiments on the use case of gantry robot scheduling, which can be considered as a continuous problem. On the one hand, the experiments indicate that both the original and the adjusted algorithm are able to learn a policy faster than a learning approach without reward machines in terms of number of actions and the number of episodes, respectively. On the other hand, the approach without reward machines is identified as data-efficient which can be considered as a factor of speed. However, it becomes clear that the original reward machine approach has difficulty learning how to behave in rare states, e.g. influenced by stochastic aspects such as machine failures. In contrast, the adjusted approach is able to deal with these challenges.

Keywords - Reinforcement learning, reward machine, automata theory, gantry robot scheduling, production system