

Mein Titel

Untertitel

Bachelorarbeit

vorgelegt am 13. Februar 2023

Fakultät Wirtschaft

Studiengang Wirtschaftsinformatik

Kurs WWI2020F

von

LEON HENNE

Betreuerin in der Ausbildungsstätte: DHBW Stuttgart:

IBM Deutschland GmbH
Sophie Lang
Senior Data Scientist

Prof. Dr. Kai Holzweißig
Studiendekan Wirtschaftsinformatik

Unterschrift der Betreuerin

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
1 Einleitung	1
1.1 Problemstellung	1
1.2 Zielsetzung	2
1.3 Forschungsfrage	2
1.4 Forschungsmethodik	2
2 Diskussion des aktuellen Stands der Forschung und Praxis	3
3 Durchführung des Laborexperiments	4
4 Ergebnisse des Laborexperiments	5
5 Reflexion und Forschungsausblick	6
Anhang	7
Literaturverzeichnis	9

Abkürzungsverzeichnis

DHBW Duale Hochschule Baden-Württemberg

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

1.1 Problemstellung

„Reinforcement Learning“ ist im Bereich des maschinellen Lernens eine Herangehensweise zur Lösung von Entscheidungsproblemen.¹ Ein Software-Agent leitet dabei durchzuführende Aktionen aus seiner Umgebung ab, mit dem Ziel die kumulierte erhaltene Belohnung zu maximieren, währenddessen sich seine Umgebung durch alle Aktionen verändert.² Die Umgebungen beinhalten in ihrer einfachsten Form eine simulierte Welt, welche zu jedem Zeitschritt eine Aktion entgegennimmt, und den eigenen nächsten Zustand sowie einen Belohnungswert zurückgibt.³ Der Einsatz einer Simulationsumgebung unterstützt besonders, wenn Limitierungen bestehen, Daten in der echten Welt zu sammeln und fürs Training zu verwenden.⁴ Eine Limitierung können bspw. Sicherheitsaspekte sein, welche beim Training von Roboterarmen, oder sich autonom bewegenden Systemen auftreten, da die einzelnen physischen Bewegungen nicht vorhersehbar abschätzbar sind.⁵ Simulationen nehmen damit als Testumgebung eine wichtige Rolle ein in der Entwicklung von Kontrollalgorithmen.⁶ Insgesamt bedarf die erfolgreiche Anwendung von Reinforcement Learning demnach nicht nur effiziente Algorithmen, sondern auch geeignete Simulationsumgebungen.⁷ Besonders schwierig, und daher sehr wichtig zu erforschen, ist es die Trainingsumgebung bestmöglich an die echte Welt anzupassen, sodass die Agenten für Roboter und autonome Fahrzeuge nach dem Training, mit generalisierten Policies in der Realität eingesetzt werden können.⁸ In der Forschungsliteratur wird diese beschriebene Problematik als „Sim to real“-Transfer beschrieben.⁹ Eine Methodik diesen Transfer zu begünstigen ist die Veränderung von visuellen oder dynamischen Parametern der Umgebung, was in der Forschungsliteratur als „Domain Randomization“ referenziert wird.¹⁰ Eine Domäne der echten Welt wird dabei eher selten durch nur eine Person oder nur eine Organisation geprägt. Oftmals beeinflussen mehrere Parteien teilweise kooperierend aber auch teilweise konkurrierend den eigenen Erfolg. Stellt man sich ein solches Szenario vor, wie bspw. einen dem Wettbewerb unterliegenden Markt, ist es naheliegend, dass jene Einflüsse auch ein Teil der Simulationsumgebung sein müssen, um ein generalisierendes Modell erlernen zu können.

Daher soll neben Domain Randomization im Rahmen dieser Arbeit untersucht werden, ob die Integration von weiteren Methoden, wie der Domain Adaptation oder dem Einsatz eines Gegen-

¹Vgl. Schuderer/Bromuri/van Eekelen 2021, S. 3

²Vgl. Schuderer/Bromuri/van Eekelen 2021, S. 3

³Vgl. Reda/Tao/van de Panne 2020, S. 1

⁴Vgl. Zhao/Queralta/Westerlund 2020, S. 737

⁵Vgl. Zhao/Queralta/Westerlund 2020, S. 738

⁶Vgl. Cutler/Walsh/How 2014, S. 2

⁷Vgl. Reda/Tao/van de Panne 2020, S. 8

⁸Vgl. Slaoui u. a. 2019, S. 1

⁹Vgl. Zhao/Queralta/Westerlund 2020, S. 738

¹⁰Vgl. Slaoui u. a. 2019, S. 2

spielers in der Simulation, die Umgebung so beeinflussen kann, dass die erlernte Policy robuster gegenüber veränderten dynamischen Bedingungen wird.

1.2 Zielsetzung

Untersuchung der erreichten Generalisierung von erlernten Policies durch Hinzufügen von Agenten als regelbasierten und als mittels RL angelernten Gegenspieler in der Simulationsumgebung. Vergleich mit bekannten Methoden der Domain Adaptation oder Domain Randomization, wie dem Verändern von Dynamikparametern der Umgebung.

Dazu

Entwicklung einer beispielhaften Simulationsumgebung zweier sich konkurrierender Spieler in Form von Flugobjekten, welche sich spielerisch gegenseitig bekämpfen. Trainieren einer Policy gegen einen Regelbasierten Gegenspieler. Training einer weiteren Policy zum Einsatz als Gegenspieler. Training einer Policy im Konkurrenzverhalten gegen den zuvor trainierten Gegenspieler. Vergleich der Policies aus dem Training mit regelbasiertem und RL gesteuerten Gegenspieler in einer Dynamik veränderten Umgebung. Testen von Performance-Metriken, wie der durchschnittlich erzielten Belohnung und Stabilitäts-Metriken, wie der Belohnungsvarianz oder der Anzahl an Abstürzen.

1.3 Forschungsfrage

Forschungsfrage: Kann durch den Einsatz von Domain Adaptation Methoden oder mittels RL trainierter Gegenspieler die Robustheit der gelernten Policy gegenüber neuen Umgebungen verbessert werden?

1.4 Forschungsmethodik

Durchführung eines quantitativen Laborexperiments zur Überprüfung von Metriken der Robustheit und Generalisierung von mittels RL erlernten Policies.

2 Diskussion des aktuellen Stands der Forschung und Praxis

3 Durchführung des Laborexperiments

4 Ergebnisse des Laborexperiments

5 Reflexion und Forschungsausblick

Anhang

Anhangverzeichnis

Anhang 1	Interview Transkripte	8
Anhang 1/1	Interview Transkript: Mitarbeiter eines Unternehmens	8

Anhang 1: Interview Transkripte

Anhang 1/1: Interview Transkript: Mitarbeiter eines Unternehmens

Literaturverzeichnis

- Cutler, M./Walsh, T. J./How, J. P. (2014): Reinforcement learning with multi-fidelity simulators. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. DOI: 10.1109/icra.2014.6907423.
- Reda, D./Tao, T./van de Panne, M. (2020): Learning to Locomote: Understanding How Environment Design Matters for Deep Reinforcement Learning. In: *Motion, Interaction and Games*. Hrsg. von Daniele Reda/Tianxin Tao/Michiel van de Panne. New York, NY, USA: ACM, S. 1–10. DOI: 10.1145/3424636.3426907.
- Schuderer, A./Bromuri, S./van Eekelen, M. (2021): Sim-Env: Decoupling OpenAI Gym Environments from Simulation Models. In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, Cham, S. 390–393. DOI: 10.1007/978-3-030-85739-4_39. URL: https://link.springer.com/chapter/10.1007/978-3-030-85739-4_39.
- Slaoui, R. B./Clements, W. R./Foerster, J. N./Toth, S. (2019): Robust Domain Randomization for Reinforcement Learning. In: *CoRR* abs/1910.10537. arXiv: 1910.10537. URL: <http://arxiv.org/abs/1910.10537>.
- Zhao, W./Queralta, J. P./Westerlund, T. (2020): Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE. DOI: 10.1109/ssci47803.2020.9308468.

Erklärung

Ich versichere hiermit, dass ich meine Bachelorarbeit mit dem Thema: *Mein Titel* selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

(Ort, Datum)

(Unterschrift)