Integration von Simulation und Reinforcement Learning zur Portalrobotersteuerung

Robert Miltenberger, Horst Zisgen (Hochschule Darmstadt)
Markus Hochhaus, Jürgen Schmidtpott (SimPlan AG)
Boris Bind (Fibro Läpple Technology GmbH)

ASIM-Fachtagung: 13-15.09.2023

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

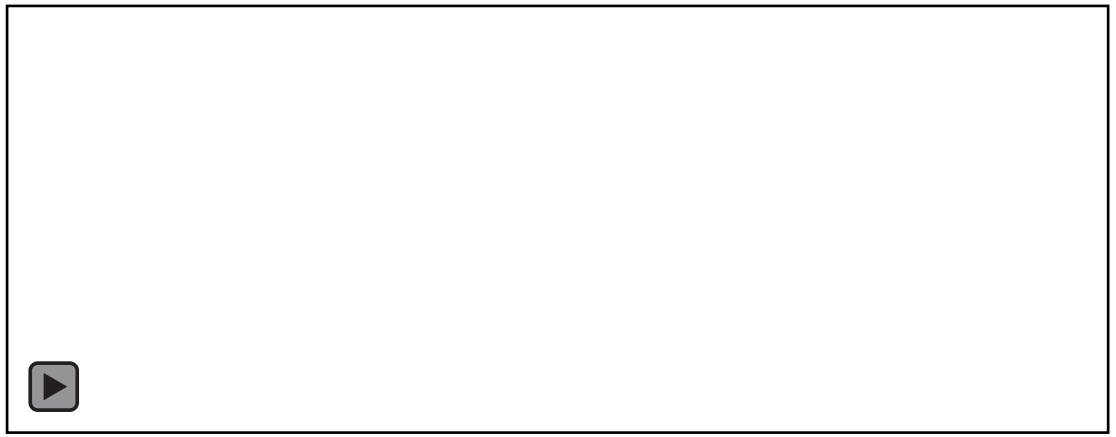
Agenda

- Motivation
- Aufbau des Simulationsmodells
- Kommunikation Reinforcement Learning (RL) Agent – Simulator
- Setup des RL-Verfahrens
- Ergebnisse
- Fazit und Ausblick

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Fragestellung

- One-Piece-Flow über verschiedene, ggf. parallele, Stationen
- Transport durch ein Portal
- Frage: Wie sieht optimale Steuerung des Portals aus?
- Status quo: Nutzung von Heuristiken



Motivation

Heuristiken …

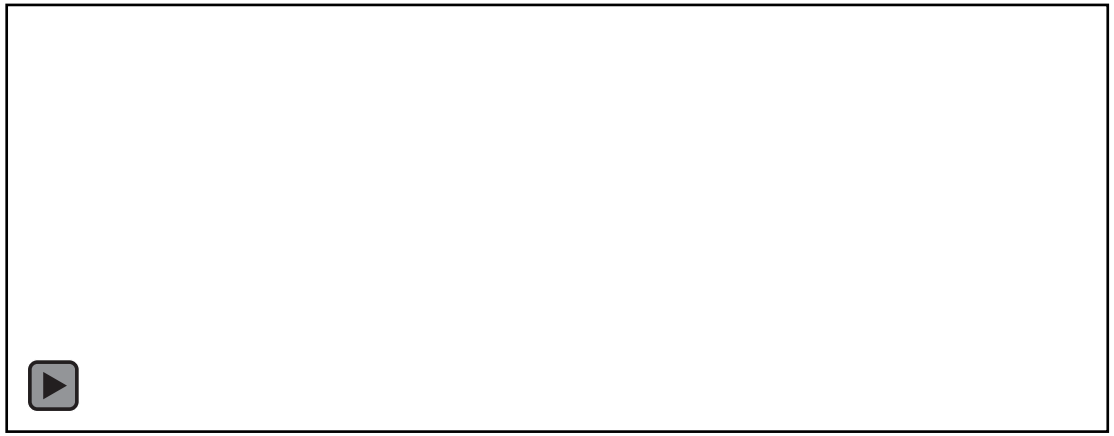
- sind immer anlagenspezifisch
- sind oft zeitaufwendig in der Erstellung
- bedürfen manueller Anpassung bei Änderungen der Anlage
- sind weniger effizient bei Anlagen mit vielen zufälligen Einflüssen (Maschinenausfälle etc.)

Idee: Nutzung von Reinforcement Learning Agenten in Kombination mit Simulation

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Allgemeiner Modellaufbau (Beispiel Plant Simulation)

- Transport durch ein Portal
  - Nutzung des Gantry-Loader-Bausteins aus der CranesAndMore-Bibliothek
- Definition von Aktionen, die an die einzelnen Loader/Greifer übergeben werden können
  - load
  - unload
  - Angabe einer Zielstation für Fahrt
  - wait
- Aufruf der Dispo der Lader bei jeder Zustandsänderung
- Allgemein gehaltene Dispo-Steuerung (Heuristik) zur Erzeugung von Referenzwerten für den RL-Agenten

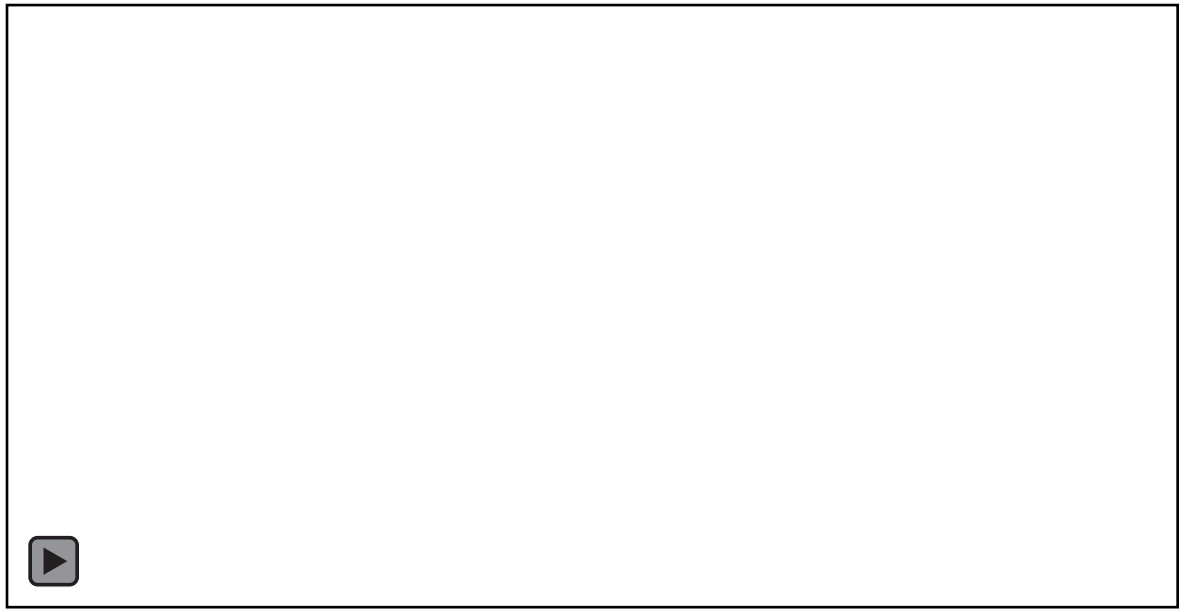


Durchlauf 3-Schritt-Modell mit I-Lader mit Störungen;
Nutzung der programmierten Nearest Job First (NJF) Heuristik

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Generisches Modell der Simulation

- Modellstruktur wird in JSON definiert
- Anlegen verschiedener Klassen in Plant Simulation (z.B. Station, Portal) und Definition von Parametrierungsmöglichkeiten
  - Mapping der Parameter auf (ggf. benutzerdefinierte) Attribute über Subtabellen
- Bei Übergabe der JSON-Datei werden angegebene Objekte erzeugt und entsprechend parametriert (Geschwindigkeit, Produktmix, Bearbeitungszeit, Stationsreihenfolge)

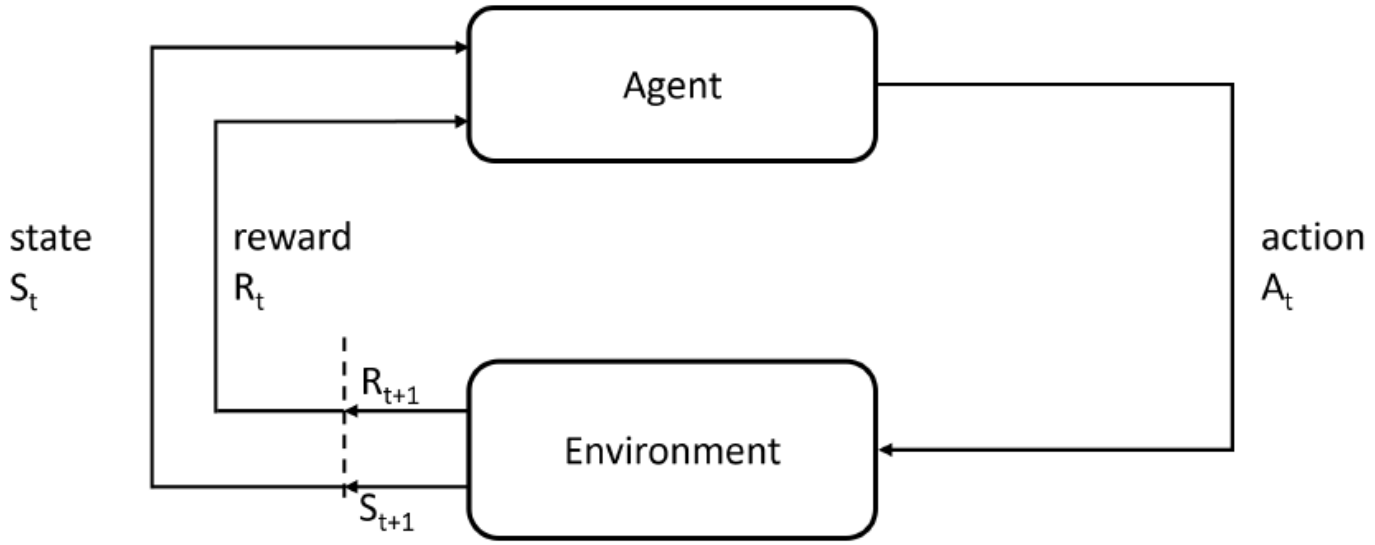


→ Beliebige Konfiguration möglich ohne Erfahrung in den entsprechenden Simulationstools

→ Änderungen am Modellverhalten immer automatisch in allen Modellen

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Reinforcement Learning



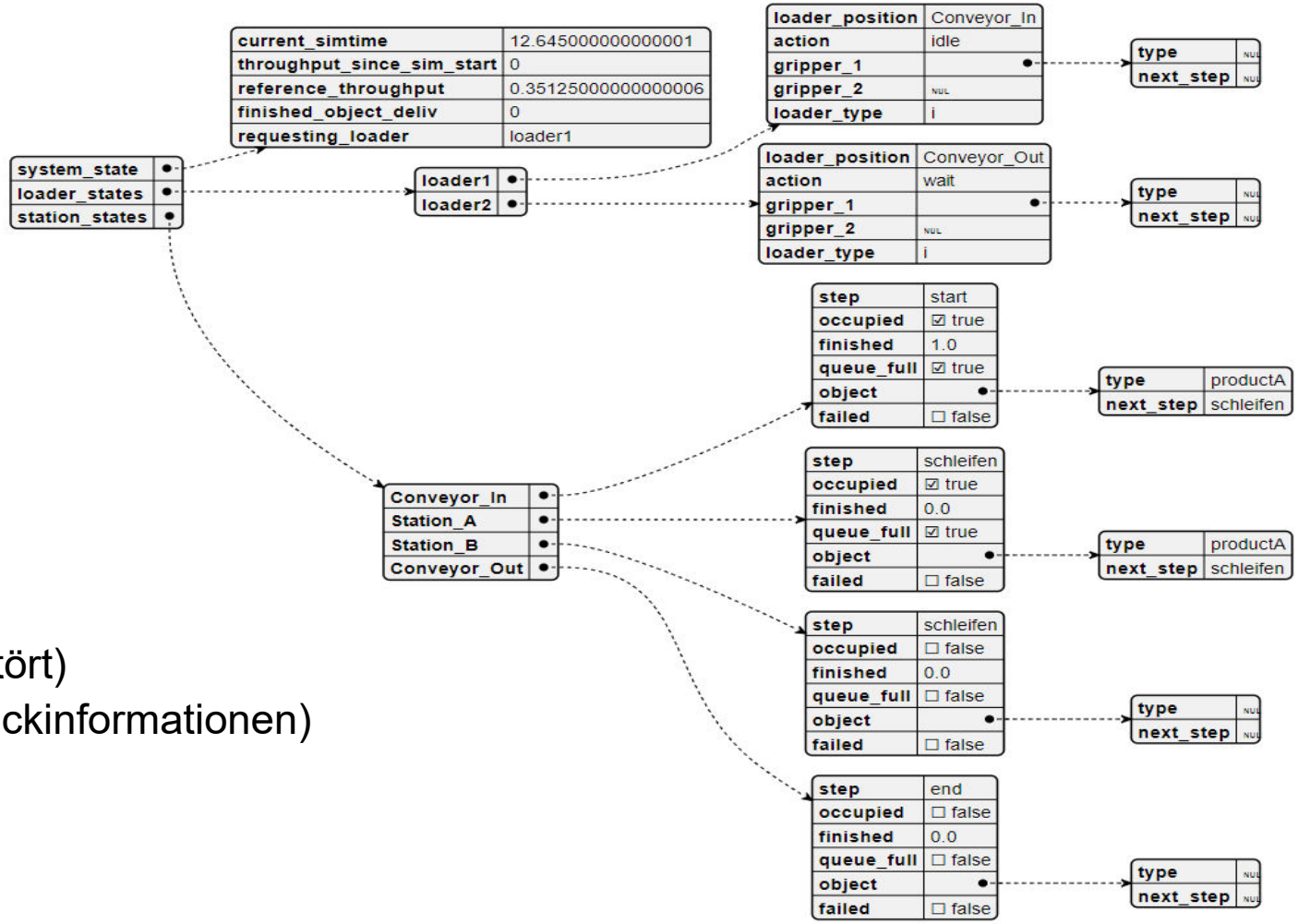
- Lernen durch Interaktion eines Agenten mit seinem Environment
- Iteration von Zustand S_t → Aktion A_t → Reward R_{t+1} → Nachfolgezustand S_{t+1} → A_{t+1} → …
- Bewertung der gewählten Aktion in einem Zustand durch sog. Q-Werte

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Query Struktur - Exemplarisch

Modellzustand

- Systemzustand
  - Simulationszeit
  - Durchsatz
- Laderzustände
  - Position
  - Ladezustand
  - Werkstückzustand
- Stationszustände
  - Prozessschritt
  - Aktueller Zustand (Belegt/leer, gestört)
  - Werkstückzustand (Fertig, Werkstückinformationen)
- Werkstücke
  - Art
  - Nächster Prozessschritt



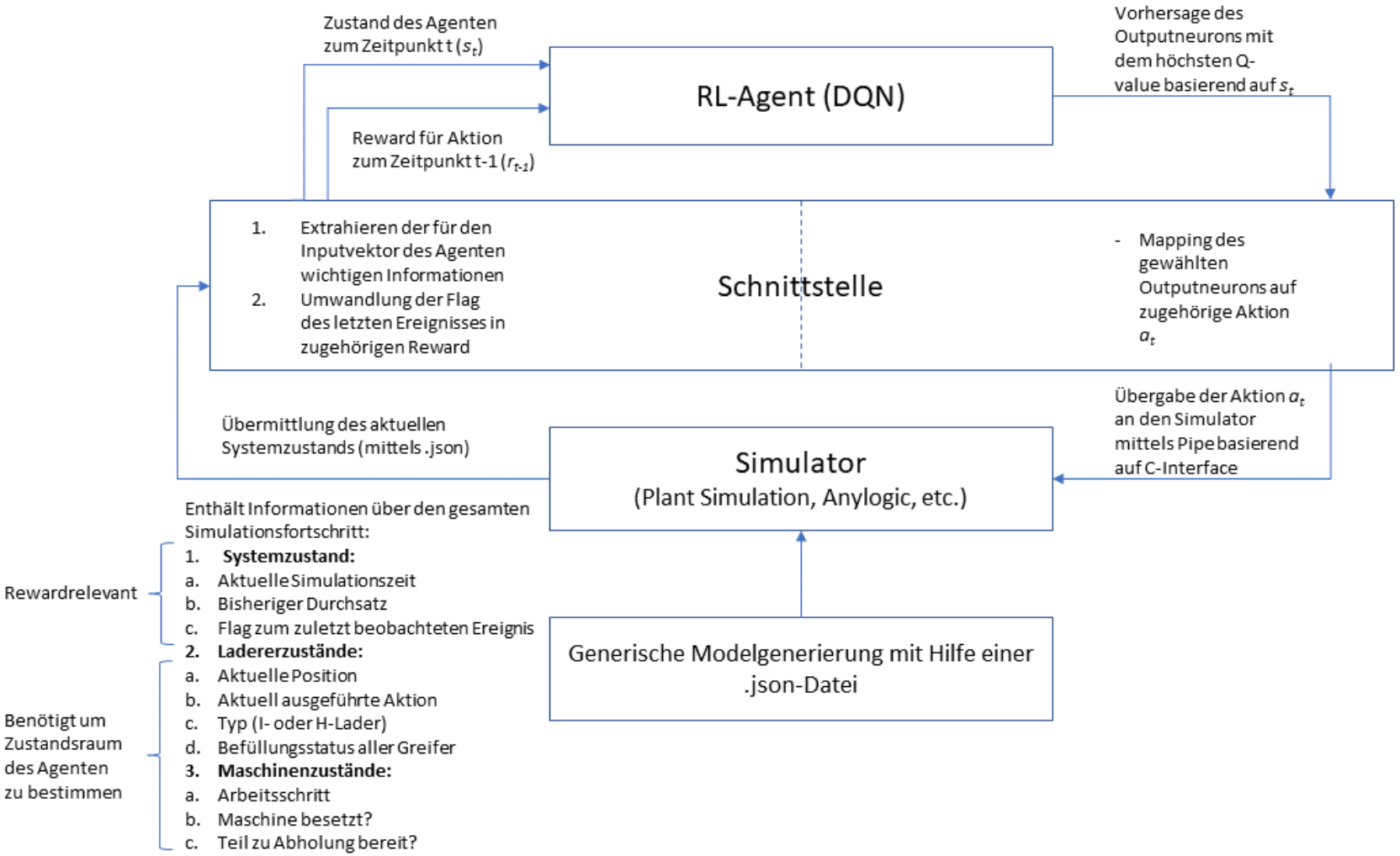
Schnittstelle

- Basierend auf HTTP (Plant: postStringToUrl / AnyLogic: http-java-modul…) und JSON
  - Simulator muss
    - http-requests erzeugen können („POST“ requests)
    - JSON erzeugen / parsen können
- Es ist egal, wie Simulator gestartet wird, aktuell via Kommandozeile

- Nach Start:
  1. Simulator meldet sich bei Schnittstelle mit seiner ID (muss beim Start übergeben werden)
  2. Setup → Fragt nach Modell und Rahmenparameter der Simulation (wie z.B. Simulationsdauer)
  3. Simulator startet die Simulation → simStart an RL-Agenten
  4. Dispo in der Simulation → Query an RL-Agent → RL-Agent antwortet mit action(s)
  5. Simulationsende → simEnd an RL-Agent → RL-Agent antwortet mit *shouldContinue* und neuem Seed
     a. *shouldContinue* = true → Sprung zu 3.
     b. ansonsten → Simulator beendet sich

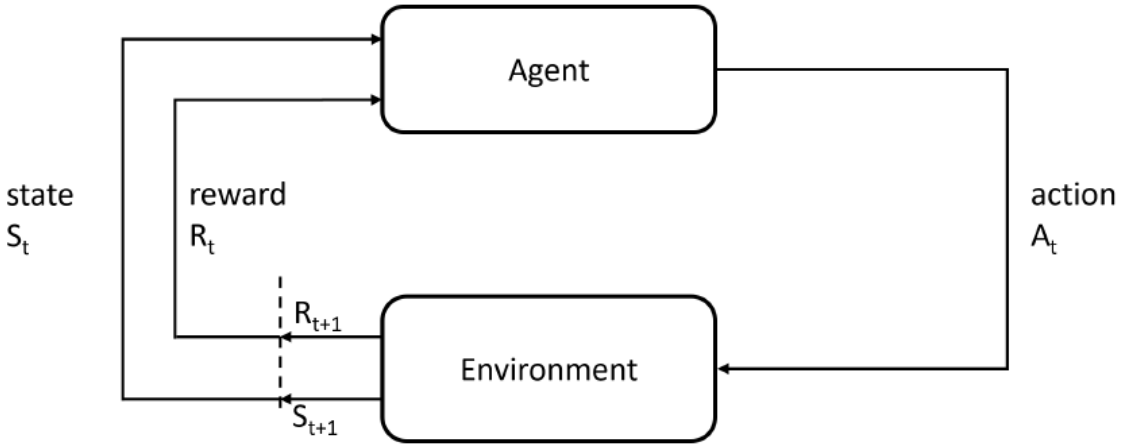
LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

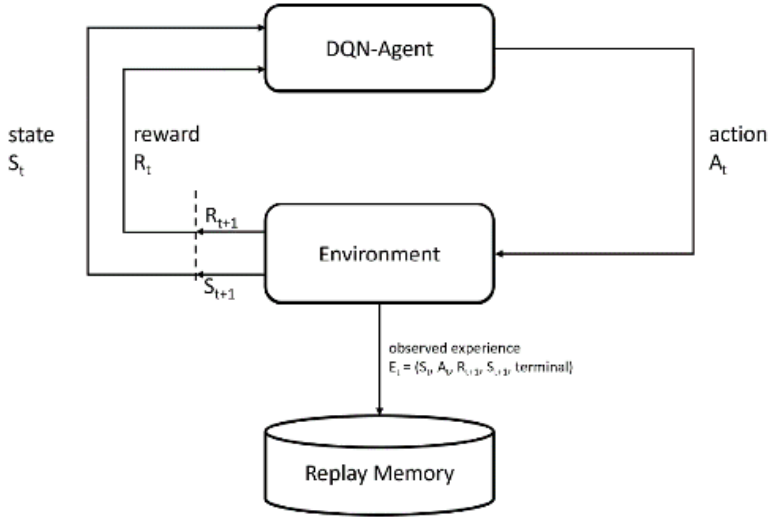
Kommunikation zwischen RL-Agent und Simulation



LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Reinforcement Learning Verfahren





- Klassisches Q-Learning iteriert den Q-Wert mittels der Formel

$$Q^{new}(S_t, A_t) \leftarrow (1 - \alpha) * Q(S_t, A_t) + \alpha(R_{t+1} + \gamma * \max_A Q(S_{t+1}, A))$$

- Q-Learning nicht praktikabel für große Zustands- und/oder Aktionsräume
- Verwendung von Deep Neural Networks (DQN) zur Approximation des Q-Werts

$$L_t = (Q(S_t, A_t; \theta) - (R_{t+1} + \gamma * \max_A Q(S_{t+1}, A; \theta^-)))^2$$

Deep Reinforcement Learning Verfahren

- Informationen des Laders und der Maschinen als Input
  - Lader: Aktuelle Position, Befüllungsstatus (0=leer, 1=unbearbeitet, 2=bearbeitet)
  - Maschinen: Teil in Bearbeitung, Teil zur Abholung bereit, Maschine ausgefallen
- Mögliche Aktionen als Output (Fahrbefehl an jede Station, Be- und Entladen und Warten)
- Unmögliche Aktionen werden unterdrückt
  - Bspw. Entladen, wenn Lader unbefüllt, etc.
- Rewardvergabe:
  - +5: fertiges Teil abgeliefert
  - +2: Teil korrekt an Bearbeitungsmaschine
  - -1: Für ersten von zwei aufeinanderfolgenden Fahrbefehlen
  - $+\frac{\tau_t}{t}$: Für Lade- und Entladebefehle, mit τ_t ist der Durchsatz bis zum Zeitpunkt t
  - 0: sonst

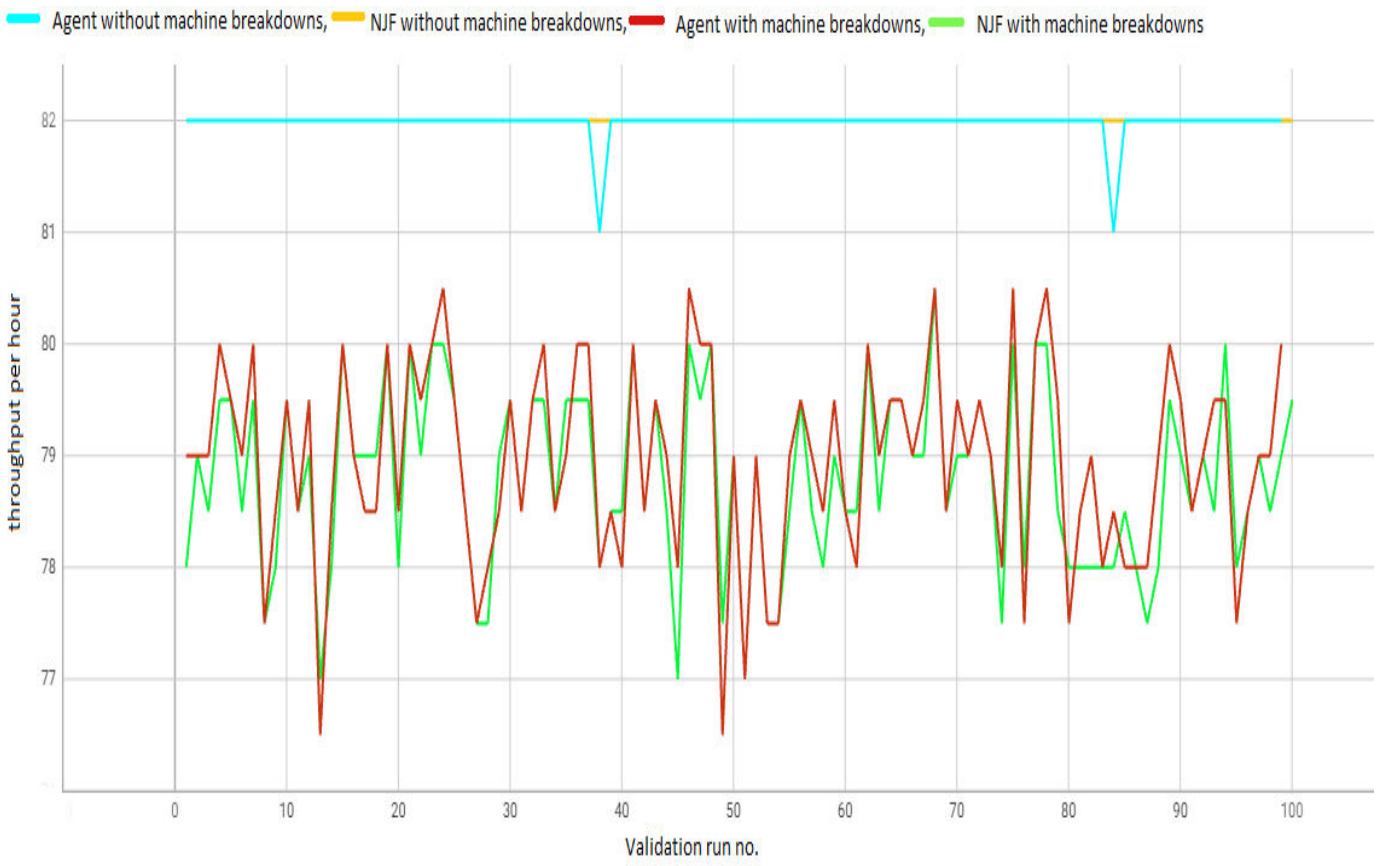
LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Ergebnisse (ohne und mit Maschinenausfällen)

„Fast“ deterministisch →

Maschinenausfälle →

	Mittlerer Durchsatz	**90% Konfidenz-intervall**
RL-Agent	78.96	[78.81,79.11]
NJF	78.79	[78.66,78.93]



LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Ergebnisse (Maschinenausfälle und unterschiedliche Zwischenankunftszeiten)

	Mittlerer Durchsatz	90% Konfidenz-intervall
RL-Agent	79.06	[78,91; 79,21]
NJF	78.40	[78,22; 78,58]

	Mittlerer Durchsatz	90% Konfidenz-intervall
RL-Agent	71.10	[70,93; 71,27]
NJF	70.85	[70,67; 71,02]

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Fazit und Ausblick

- **Fazit**
  - Generischer Modellaufbau für das Zusammenspiel Simulation – RL-Agent essentiell
  - Modellierung der Bewertung (Reward-Funktion) im Simulationsmodell erfordert besondere Sorgfalt
  - Für die beobachteten Szenarien liefert der RL-Agent im Vergleich zur NJF-Heuristik mindestens gleich gute Ergebnisse

- **Ausblick**
  - Betrachtung komplexerer Modellkonstellationen (unterschiedliche Ladertypen, Mehrladermodelle, mehrere Produkttypen, Prüfkonzepte, etc.)
  - Bereits erste gute Ergebnisse für H-Lader-Portale und Portale mit mehreren Produkten
  - Entwicklung komplexerer Agentenstrukturen, bspw. Multiagenten-Systeme zur Steuerung mehrerer Lader oder Reward-Machines

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Förderung

Das KISPo Projekt (HA Projektnum. 1286/21-187) wird aus Mitteln der LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben gefördert.

Setting des Neuronalen Netzes

- Informationen des Laders und der Maschinen als Input
  - Lader: Aktuelle Position, Befüllungsstatus (0=Leer, 1=unbearbeitet, 2=bearbeitet)
  - Maschinen: Teil in Bearbeitung, Teil zur Abholung bereit, Maschine ausgefallen?
- Mögliche Aktionen als Output (Fahrbefehl an jede Station, Be- und Entladen und Warten)
- 3 Hidden Layers á 16 Neuronen
- ReLu-Aktivierungsfunktion zwischen Hidden Layern
- Parameteroptimierung mittels ADAM-Optimizer

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187

Trainingssetting

- Episoden: 20.000
- ε_{decay}: 0,99975
- Replay-Memory Größe: 50.000
- Minibatchsize: 64
- Training: Mit 1/3 Wahrscheinlichkeit nach jeder Entscheidung
- Target-Prediction Update: 10 Episoden

LOEWE Exzellente Forschung für Hessens Zukunft
HESSEN Hessisches Ministerium für Wissenschaft und Kunst
HA-Projektnr. 1286/21-187