TECHNICAL APPLICATIONS AND DATA MANAGEMENT. WS 2021/2022.

VORLESUNG 11

14.12.2021

MÜNCHEN

STUDIENGANG DIGITAL MANAGEMENT.

AGENDA

1. Status Projektarbeit AI
2. Ausblick
    1. Machine Learning in der Cloud
    2. Reinforcement Learning

WAS HABEN WIR BIS JETZT GEMACHT?

ROADMAP	WAS HABEN WIR GEMACHT?
Vorlesung 1	Workflow Data Management, Datentypen und Datenqualität
Vorlesung 2	Einführung Data Science und Data Science Workflow, Grundlagen Data Management
Vorlesung 3	Grundlagen Stochastik: Wahrscheinlichkeitsrechnung, deskriptive und explorative Statistik
Vorlesung 4	Statistische Inferenz, lineare Regression
Vorlesung 5	Einführung Machine Learning, Unüberwachtes Lernen
Vorlesung 6	Überwachtes Lernen
Vorlesung 7	Neuronale Netze und Convolutional neural networks
Vorlesung 8	Aufgabenstellung Data Science, Case Study CNN: Malaria
Vorlesung 9	Aufgabenstellung AI, RNN, Case Study RNN
Vorlesung 10	Status Projektarbeit Data Science und Fragen & Aufgabenstellung AI

1. STATUS PROJEKTARBEIT AI

2. AUSBLICK

2.1 AI IN DER CLOUD

EINLEITUNG

- Bis jetzt haben wir vorwiegend auf **einem** Rechner (lokal/ Cloud) unsere Use Cases umgesetzt.
- Wir haben gesehen, daß Machine Learning sehr ressourcen- und zeitintensiv ist:
  - Speicherbedarf: je mehr Daten, desto bessere Ergebnisse[1] (Datenmenge x 2 oder 3 besser als jedes Parameter-Tunen).
  - Rechenbedarf: je länger trainiert wird, desto (meist) besser sind die Ergebnisse.

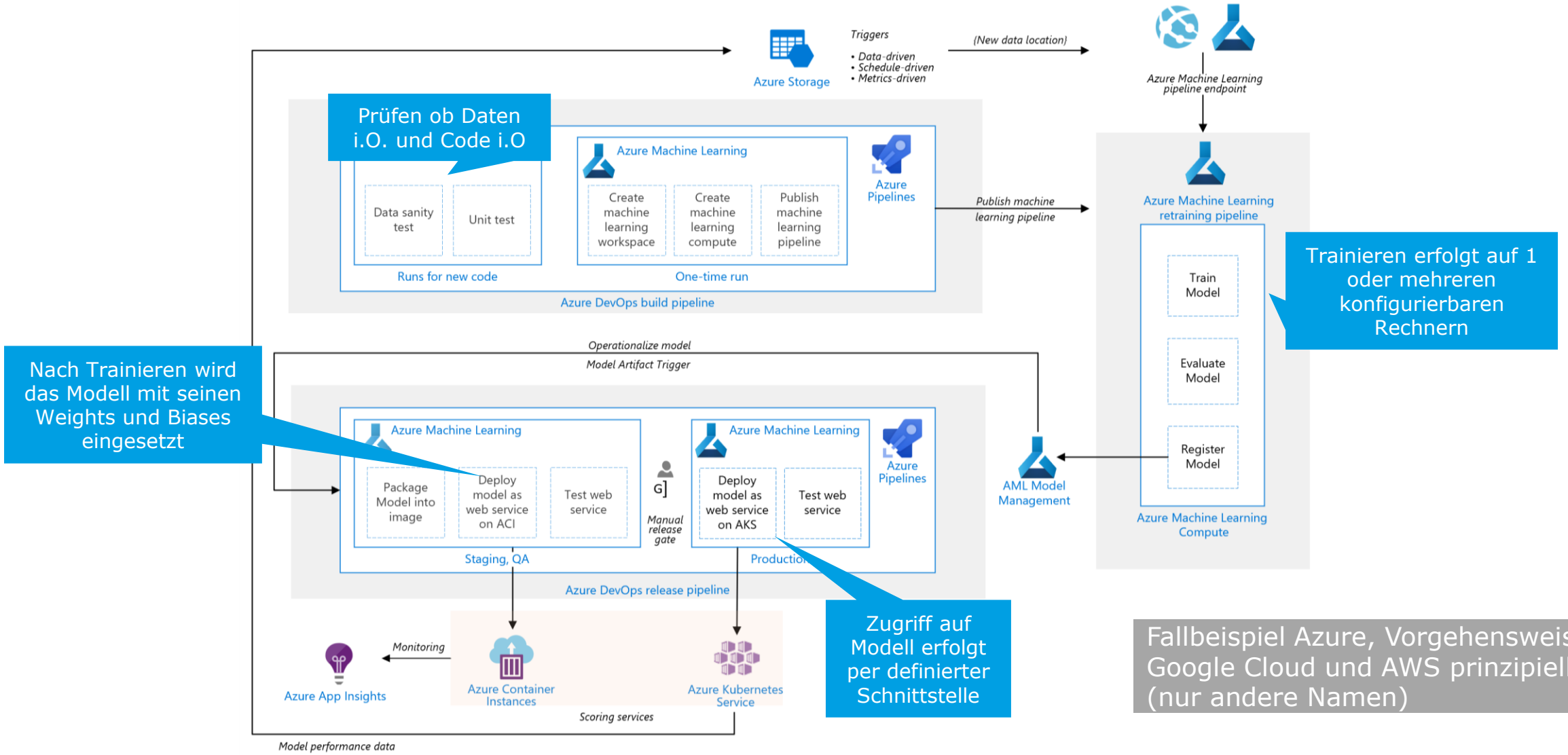
- Nach anfänglichem Training werden vor allem die Rechenbedarfe nicht mehr kontinuierlich in dem Ausmaß benötigt.

➔ Große wirtschaftliche Vorteile bei Nutzung von Cloud-Anbietern statt Aufbau eigener Strukturen (on-premise).

➔ Auch Startups können ohne große Investitionen „gleiche" Ressourcen wie große Firmen nutzen.

Quellen: 1 Banko, Brill: Scaling to very very large corpora for natural language disambiguation, 2001.

ÜBERSICHT ARCHITEKTUR MACHINE LEARNING IN DER CLOUD.



Fallbeispiel Azure, Vorgehensweise bei Google Cloud und AWS prinzipiell gleich (nur andere Namen)

Quelle: Link

AUSBLICK.

- Starker Wettbewerbsdruck zwischen Amazon Web Services (erster Anbieter!), Microsoft Azure und Google Cloud.
- Etablierte IT-Anbieter werden in das Cloud-Geschäft einsteigen: SAP, Oracle, ….
- Dieser steigende Wettbewerb führt zu:
  - Sinkenden Preisen.
  - Firmen: Unterstützungsleistungen bei Einführung der Cloud.
  - Privatanwender: freie Kontingente für Ausprobieren (200$ Azure, 300$ Google).
  - Allgemein: Entwicklung von automatisierten ML-Ansätzen (AutoML), Bereitstellen von Standardlösungen, Tutorials, …

In dieser Vorlesung haben Sie die Grundlagen gelernt; schauen Sie sich doch (bei Interesse ;-)) die Tutorials an!

2.2 REINFORCEMENT LEARNING

ÜBERSICHT/ CLUSTERING MACHINE LEARNING ALGORITHMEN.

Unsupervised Learning

Lernen **ohne** vorher definierte **Zielwerte** oder Belohnung

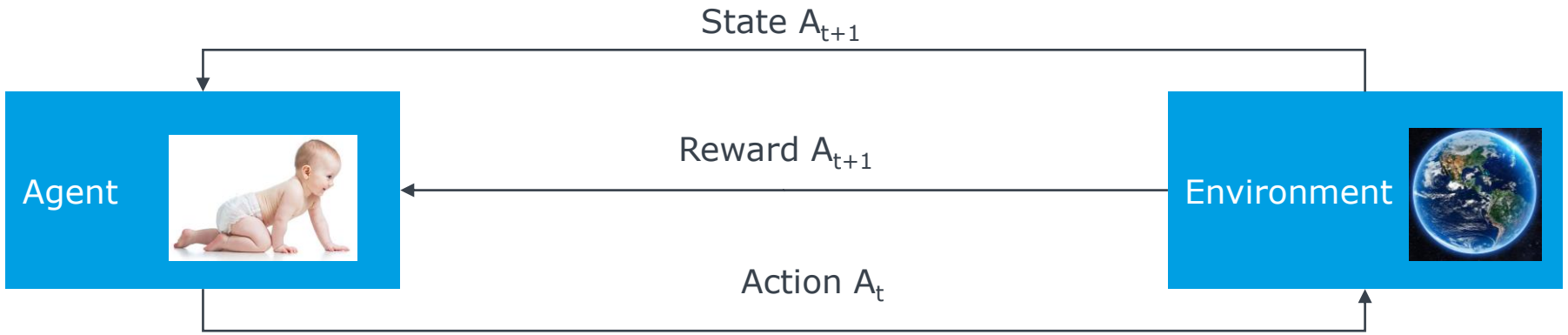
Supervised Learning

Algorithmus lernt eine Funktion, die Eingabegrößen auf **vorher** definierte Outputs mappt.

Reinforcement Learning

Agent/ Algorithmus lernt **selbständig** mit Ziel, eine Belohnung zu **maximieren**.

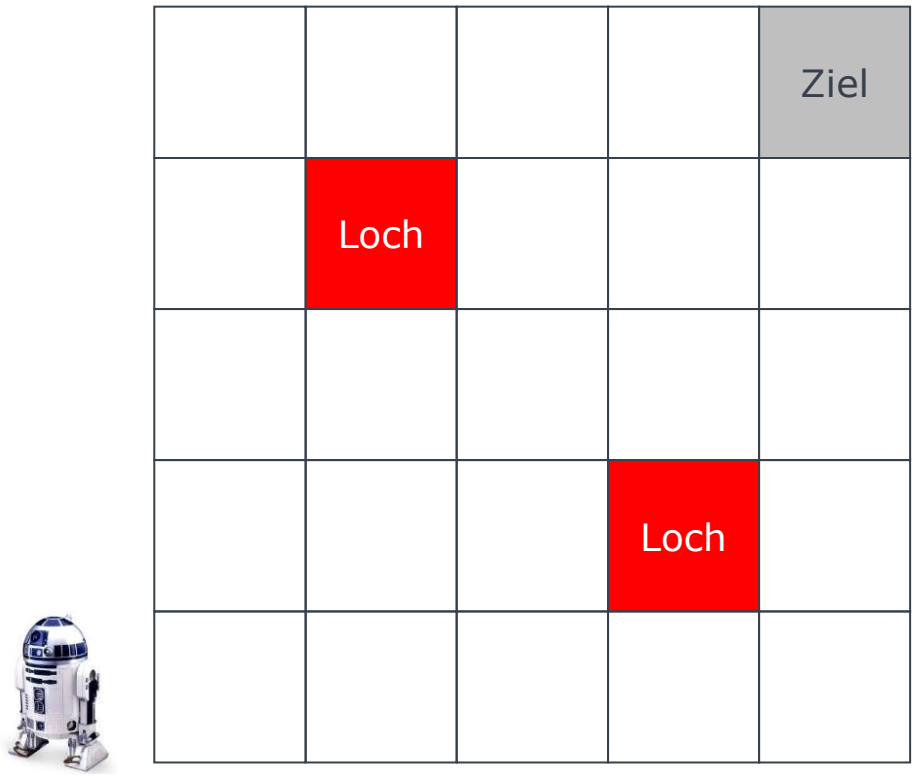
VEREINFACHTE DARSTELLUNG REINFORCEMENT LEARNING.



Unterschiede zu bisherigen Verfahren:
- Algorithmus agiert, um Belohnung zu maximieren und nimmt somit Einfluß auf seine Umgebung (und damit auf die nachfolgenden Daten).
- Agent weiß nicht (immer), wie die Umwelt auf seine Aktionen reagiert.
- Diese Belohnung erfolgt nicht zwingend per direktem Feedback, sondern auch ggf. später.
- Deshalb immer zeitabhängige, sequentielle Daten.
- Kann zur Laufzeit auch vorher nicht gelernte Sachen lernen (supervised Models sind statisch!).

Quellen: CS231n: Convolutional Neural Networks for Visual Recognition, Link
Introduction to RL with David Silver, Link

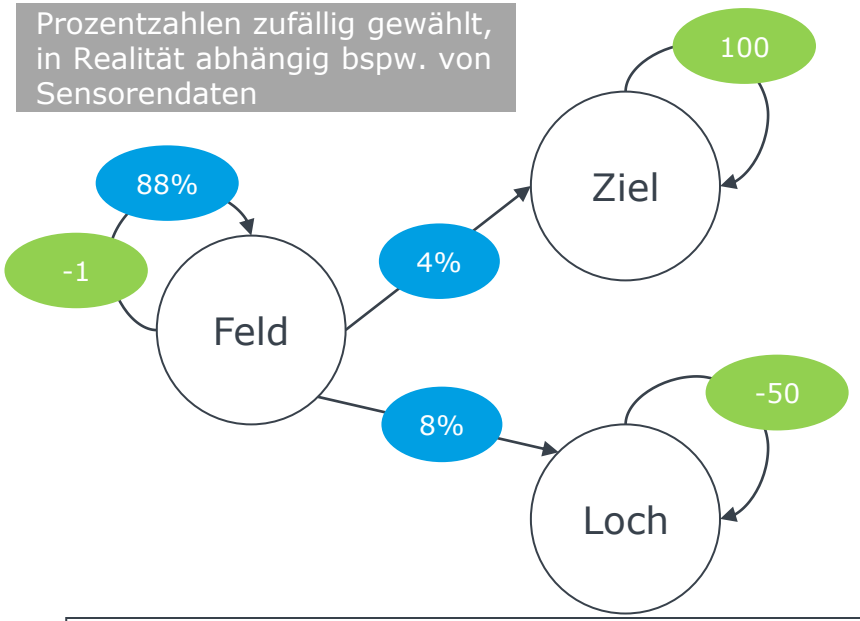
REINFORCEMENT LEARNING: FALLBEISPIEL WEGFINDEN ROBOTER.



Ziel: Gegeben zufällige Startposition Roboter, finde das Ziel mit der geringsten Anzahl von Schritten.

FORMALE GRUNDLAGE VON REINFORCEMENT LEARNING SIND MARKOV DECISION PROCESSES (MDP).

Prozentzahlen zufällig gewählt, in Realität abhängig bspw. von Sensorendaten



Eigenschaften:

- S = Menge aller möglichen Zustände/ States s
- A = Menge aller Aktionen,
- T(s, a, s`) = Übergangsfunktion von einem Zustand s in nächsten s` durch Aktion a. Diese wird in Prozent angegeben, da Unsicherheit (keine vollständige Sicht auf die Umwelt!!)
- R(s, a, s`) = Belohnung für diesen o.a. Übergang. Da wir mehrere Schritte haben können, gilt:

Aktuelle Belohnung

$$G_t = R_{t+1} + \gamma R_{t+2} + \ldots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Anschließende Belohnung inkl. Gewichtungsfaktor Wichtigkeit zukünftiger Schritte γ (1:= sehr, 0:= gar nicht)

- V(s) = Wert eines Status. $V(s) = \mathbb{E}[G|S_t = s]$

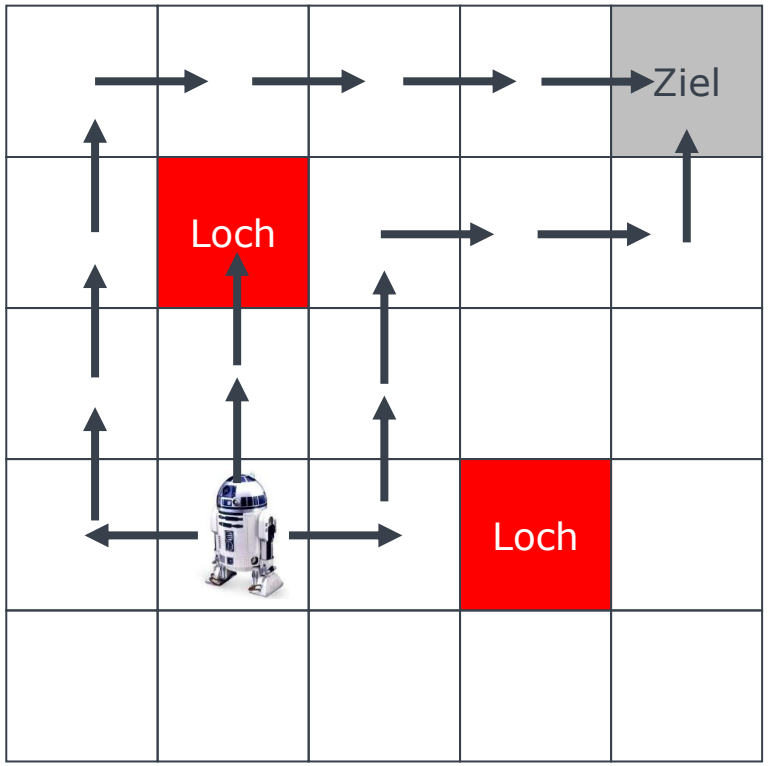
Gemittelte/ erwartete Belohnung im Zustand s

Fürs Fallbeispiel:
Zustände S:={Feld, Ziel, Loch}
Aktionen A: ={←, ↑,→,↓}
Übergangsfunktion T(Feld, *, Ziel) = 4%
Reward R(Feld, *, Ziel) = 100,
Reward R(Feld, *, Loch) = -50
Reward R(Feld, *, Feld) = -1 (Ziel: schneller Weg!)
V(Feld) = -1*0,88, V(Ziel) = 100*0,04 + -1*0,88

Vereinfachte Annahme bei Markov-Prozessen: nur der aktuelle Zustand ist relevant für die Zukunft, nicht die vorherigen!

REINFORCEMENT LEARNING: FALLBEISPIEL WEGFINDEN ROBOTER.

Welcher der 3 Wege ist der Beste?



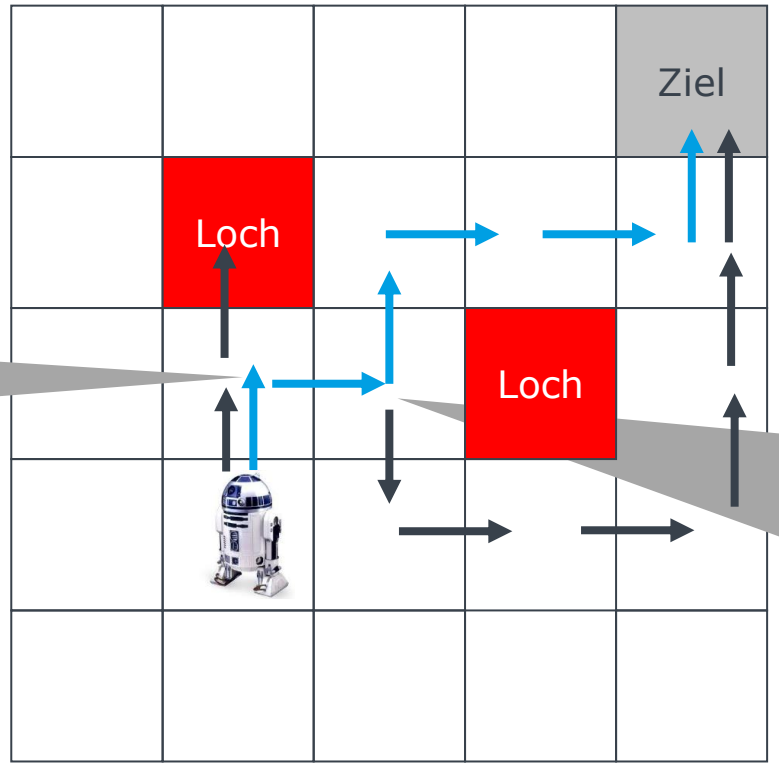
Ziel: Finde das Ziel mit der geringsten Anzahl von Schritten

Quellen: CS231n: Convolutional Neural Networks for Visual Recognition, Link
Introduction to RL with David Silver, Link

REINFORCEMENT LEARNING: FALLBEISPIEL WEGFINDEN ROBOTER.

Wie würde ein Mensch das lösen?
Stets prüfen, was die beste Aktion ist

Entscheidung RL Agent:
Weg 1: führt ins Loch, R = -50
Weg 2: kein Loch



Entscheidung RL Agent:
Blaue Weg: in weniger Schritten zum Ziel
Schwarze Weg: mehr Schritte zum Ziel.

Wir hatten für eine schnelle Zielfindung definiert, daß Schritte die nicht direkt zum Ziel führen, eine Belohnung von -1 erhalten.

Dadurch erhält der blaue, kürzere Weg eine höhere Belohnung als der schwarze und wird deshalb gewählt.

Aber wie trifft und lernt eine Maschine solche Eigenschaften?

Quellen: CS231n: Convolutional Neural Networks for Visual Recognition, Link
Introduction to RL with David Silver, Link

FUNDAMENTALE PROBLEM DES REINFORCEMENT LEARNING: EXPLORATION VS. EXPLOITATION

Exploration:

- Initiale Umgebung & Auswirkungen auf diese unbekannt
- Umwelt kann sich über Zeit ändern
- Agent muß explorieren, um Umwelt kennenzulernen. Das heißt, er führt eine Aktion aus, erfaßt die geänderte Umwelt und die Belohnung und speichert das ab.

Exploitation

- Ziel des Agenten ist Maximieren Belohnung.
- Agent führt also die Aktion bzw. Sequenz von Aktionen aus, die ihm am meisten Belohnung zurückgibt.

Dilemma: Wenn der Agent nur Exploitation wählt, lernt er nie, ob es nicht bessere Aktionen gibt.
Wenn der Agent nur exploriert, wird der Nutzen nicht maximiert und viele schlechte Aktionen ausgeführt.
➔ (langfristiger) Ausgleich notwendig.

Falls Sie immer in die gleichen Bars, Clubs, Restaurants gehen, lernen Sie nie besseres kennen.
Andererseits vermeiden Sie so Reinfälle.
Aber irgendwie müssen Sie die für Sie besten Läden ja auch mal kennengelernt haben….

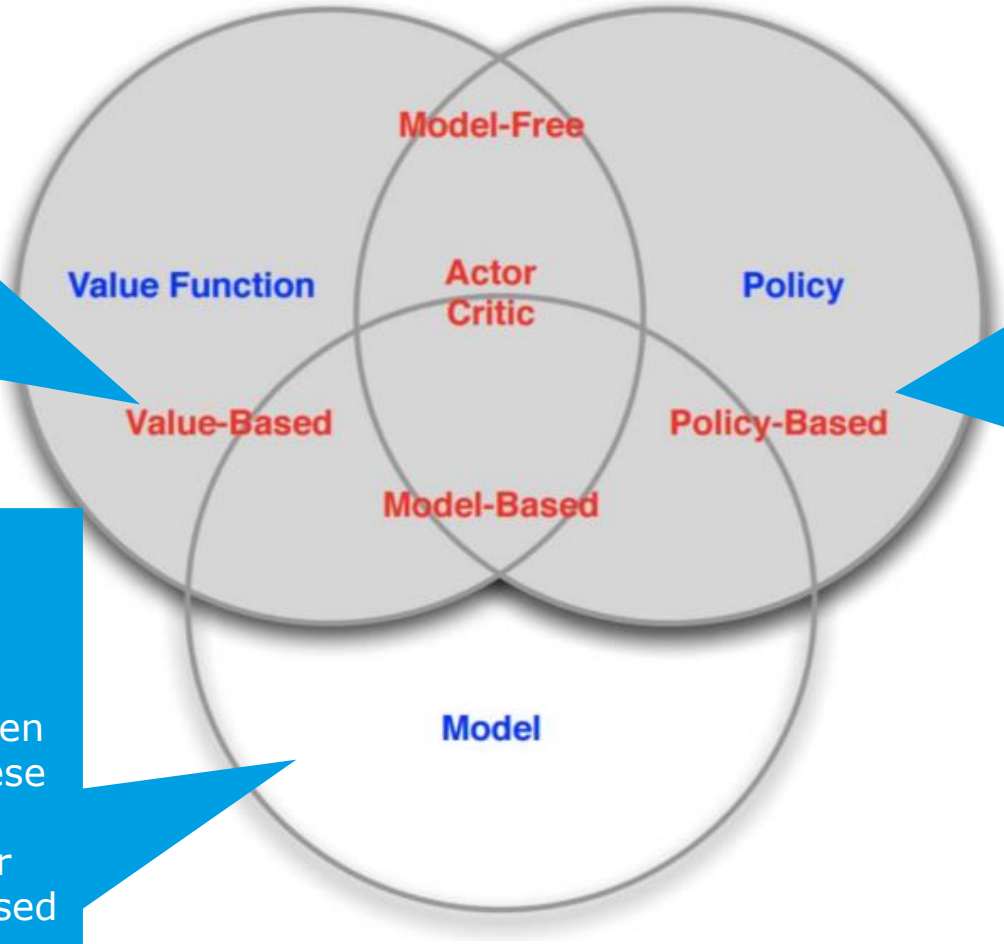
Lösung Ausgleich Exploration vs. Exploitation durch Epsilon-Greedy[1]-Verfahren:

- Initiale Definition eines Wertes ε, bspw. 0,1.
- Es wird vor jeder Aktion eine Zufallszahl p berechnet.
- Falls $p < \varepsilon$ dann beliebige Aktion ausführen, sonst beste Aktion (mit Wahrscheinlichkeit $1 - \varepsilon$).
- In der Praxis starten wir mit einem höheren ε-Wert, der dann kontinuierlich verringert wird.

Quelle: Sutton, R & Barto, A.: "Reinforcement learning: an introduction", 1998.

ÜBERSICHT ENTSCHEIDUNGSVERFAHREN FÜR RL-AGENTEN.

- Iterative Berechnung der Value-Funktion mit höchstem Wert, dann indirektes Ableiten der Entscheidung für Aktion (Policy) aus dieser Value-Fkt. durch Wahl des Status mit höchstem Wert
- Vorteil: geringer Speicherbedarf, keine Tabelle notwendig.
- Nachteil: Aktionen nicht direkt ableitbar.



- Algorithmus lernt Tabelle mit Zusammenhang Status s und Aktion a.
- In jedem Zustand s wird aus der Tabelle die Aktion a gewählt, die die höchste Belohnung verspricht
- Vorteil: Aktionen direkt ableitbar
- Nachteil: sehr hoher Speicherbedarf für Tabelle bei komplexen Problemen

- Agent lernt Modell, wie Umwelt funktioniert basierend auf seinen Aktionen und plant darauf basierend seine Aktionen.
- Vorteil: kann auch mit sehr komplexen Umgebungen umgeben und lernt diese schneller.
- Nachteil: Modell kann schwer lernbar sein (und dann ist simples Policy-Based einfacher).

Detaillierung im folgenden

Quellen: Introduction to RL with David Silver, Link

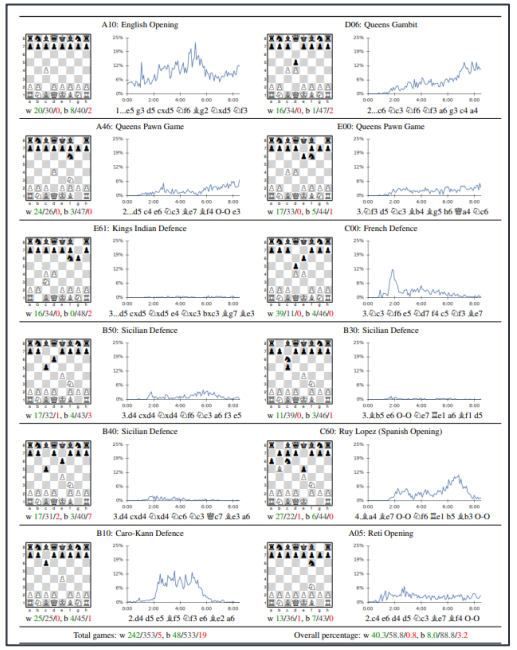
FALLBEISPIELE MODEL-BASED REINFORCEMENT LEARNING.

Mnih et al.: “Playing Atari with Deep Reinforcement Learning”, 2013.

→ Erster Einsatz von Deep Learning Verfahren für Lernen des Modells (aus Sensorendaten).

Silver et al.: “Mastering the game of Go with Deep Neural Networks & Tree Search”, 2016.

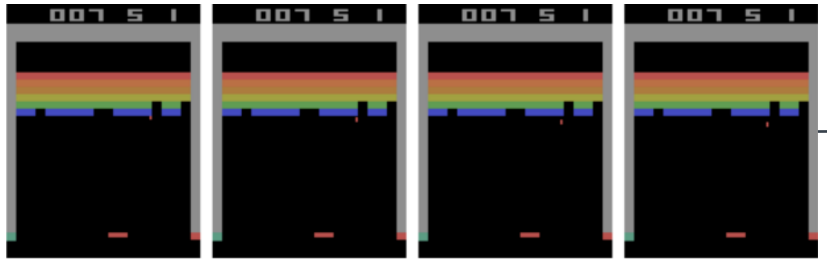
→ Go galt aufgrund seines gigantischen Zustandsraums als nicht lernbar für Computer

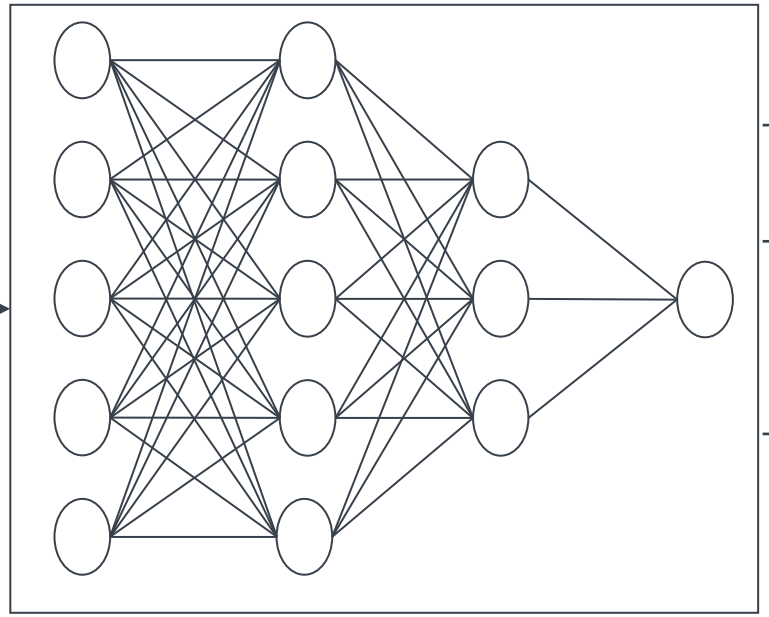


Silver et al: “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm”, 2017.

→ Weiterentwicklung Go-Ansatz, Modell lernt komplett eigenständig.

WIE FUNKTIONIERT MODEL-BASED RL? (VEREINFACHT)





Q-value Action 1

Q-value Action 2

...

Q-value Action n

Wir nehmen 4 sequentielle Frames, um einen Zeitverlauf abzubilden

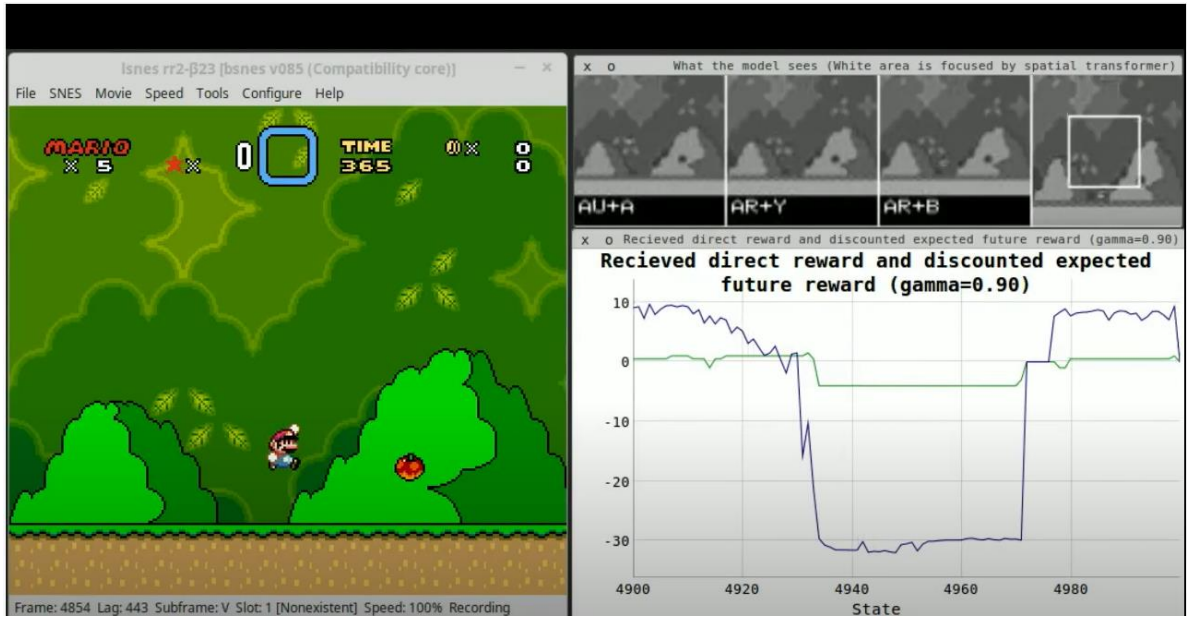
CNN reduziert die Bilder auf die wesentlichen Inhalte

Wähle das State-Action Paar mit höchstem Wert

Formal: $Q(s_t, a_t) \leftarrow (1 - \alpha) * \underbrace{Q(s_t, a_t)}_{alter\ Wert} + \alpha * \overbrace{(r_t + \gamma * \underbrace{max_a Q(s_{t+1}, a))}_{gesch\ddot{a}tzter\ zuk\ddot{u}nftiger\ optimaler\ Wert}}^{Gelernter\ Wert}$ (Bellman-Gleichung)

Quelle: basierend auf Link

LIVE DEMOS.



https://www.youtube.com/watch?v=V1eYniJ0Rnk

https://www.youtube.com/watch?v=L4KBBAwF_bE

LITERATUR UND WEITERE QUELLEN (AUSZUG).

Künstliche Intelligenz:

- Russel, Norvig: Artificial Intelligence – a modern approach
- Lapan: Deep Reinforcement Learning Hands-on.
- Silver: Introduction to Reinforcement Learning (Link)
- Barto, Sutton: Reinforcement Learning (Link)

Online-Kurse (bei Interesse):

- Kostenfrei: Udacity: Reinforcement Learning, Link
- Coursera: Reinforcement Learning, Link
- Udacity: Deep Reinforcement Learning, Link