Reinforcement Learning in der Modellfahrzeugnavigation

von

Manuel Trittel

Informatik

HAW Hamburg

Vortrag im Rahmen der Veranstaltung AW2 im Masterstudiengang, 02.07.2009

Einführung

Einführung
- Thema
- Problemstellung

Relevante Reinforcement Learning Verfahren
- Dynamische Programmierung
- Temporal Difference Learning
- Generalisierung und Funktionsapproximation

Anwendungsfälle der Verfahren

Praktische Herausforderungen
- Beschleunigung der Lernverfahren
- Diskretisierung kont. Zustands- und Aktionsräume

Zusammenfassung

Bedeutung für eigene Arbeitsziele

Gliederung

Einführung

Thema

Reinforcement Learning
in der Modellfahrzeugnavigation

Konkreter Anwendungsfall:

Geschwindigkeitsregelung

Geschwindigkeitsmaximierung
= Zeitminimierung

Einhaltung einer maximalen Zentripetalkraft

Gliederung

Einführung

Problemstellung

Gliederung

Relevante Reinforcement Learning Verfahren

Dynamische Programmierung

Temporal Difference Learning

Generalisierung und Funktionsapproximation

Gliederung

Einführung
- Thema
- Problemstellung

Verfahren
- Dyn. Programmierung
- TD Learning
- Generalisierung u. FA

Anwendungsfälle

Herausforderungen
- Beschleunigung
- Diskretisierung

Zusammenfassung

Bedeutung f. AZ

Relevante Reinforcement Learning Verfahren

Dynamische Programmierung

aus den 1950er Jahren von Richard E. Bellman

basiert auf vollständigem Markov Modell

Bestimmung einer optimalen Zustandstrajektorie durch

value iteration

policy iteration

$$V_k(s) = \max_a (R(s,a) + \gamma \cdot V_{k-1}(s'))$$

-7	-6	-5	-6	-7
0	-5	-4	-5	-6
-1	-2	-3	-8	-7

Gliederung

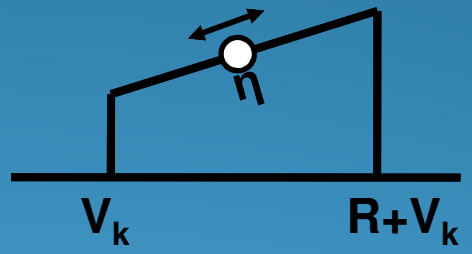
Relevante Reinforcement Learning Verfahren

Temporal Difference Learning (TD Learning)

schrittweise Update der *value function*

Zustandswert wird beim Verlassen geupdated:

$$V_{k+1}(s_t) = (1-\eta) \cdot V_k(s_t) + \eta(R(s,a) + \gamma \cdot V_k(s_{t+1}))$$



$\Delta V(s_t)$ wird als *temporal difference* bezeichnet

prinzipiell analog mit der Q-Funktion

$$td = R(s,a) + \gamma \cdot Q^{\pi}(s',a') - Q^{\pi}(s,a)$$

mit π als *estimation policy*

Gliederung

Relevante Reinforcement Learning Verfahren

Temporal Difference Learning (TD Learning)

Strategien...
zur praktischen Anwendung
zum Erlernen der Q-Funktion

eine Strategie für beides Nutzen
***on-policy* Verfahren (SARSA)**

jeweils eine unabhängige Strategie
***off-policy* Verfahren (Q-learning)**

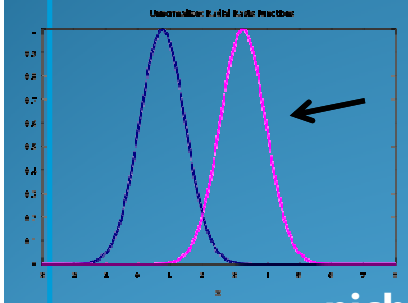
Gliederung

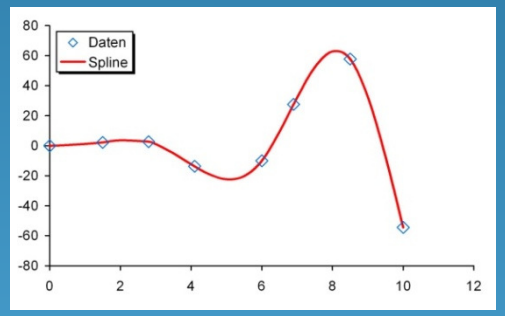
Relevante Reinforcement Learning Verfahren

Generalisierung und Funktionsapproximation

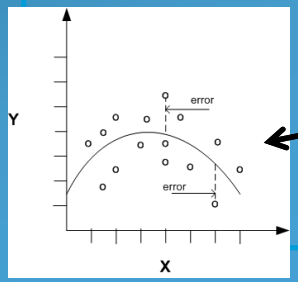
Problem: Sehr große Zustands- u. Aktionsräume
- **zu wenig Speicherplatz**
- **zu wenig Rechenzeit**
- **zu viel Zeit zum Lernen benötigt**

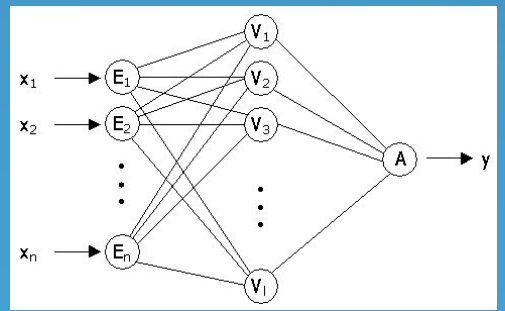
Funktionsapproximation nach [Sutton u. Barto]:
- **Gradienten-Abstiegs-Methoden**
- **lineare Methoden**
  - **Grob- u. Teilkodierung**
  - **Kubische Splines** →
  - **Radiale Basisfunktionen**





- **nicht lineare Methoden**
  - **Mustererkennung**
  - **neuronale Netze** →
  - **Regressionsmethoden**





Gliederung

Einführung
- Thema
- Problemstellung

Verfahren
- Dyn. Programmierung
- TD Learning
- **Generalisierung u. FA**

Anwendungsfälle

Herausforderungen
- Beschleunigung
- Diskretisierung

Zusammenfassung

Bedeutung f. AZ

Anwendungsfälle

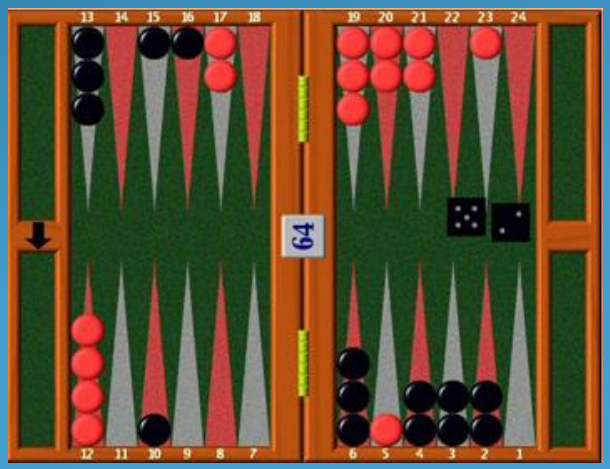
Anwendungsfälle aus [Kaelbling u.a.]

Devilsticking robot
DP

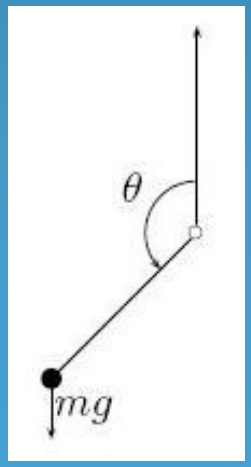
Boxpushing robots [Mahadevan]
(Q-learning)

Fahrstuhlsteuerung [Crites u. Barto]
(Q-learning, MLP)

Backgammon [Tesauro]
MLP mit TD(λ)



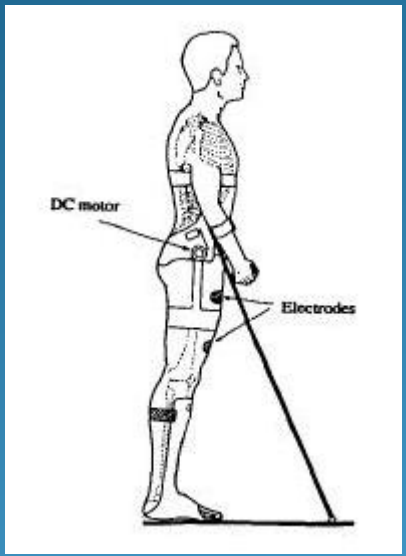
2-gliedriges Pendel [Coulom]
TD(λ) mit MLP / Gauss network

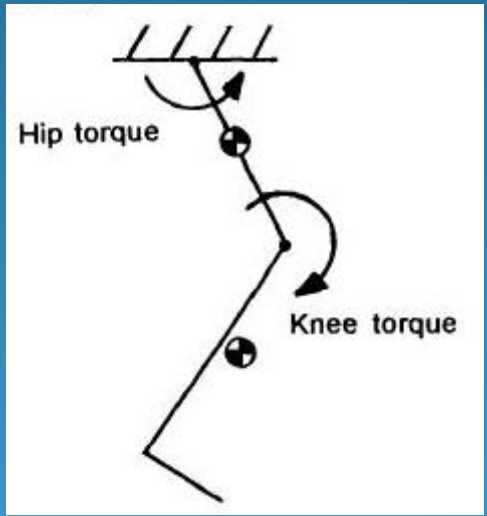


Gliederung

Anwendungsfälle

Prothese (Bein/Hüfte/Knie) aus [Thrasher]
SL + RL [Williams] Reinforce Algorithmus





Gliederung

Praktische Herausforderungen

Beschleunigung der Lernverfahren
- **Vorwissen über die Systemumgebung**
- **Eligibility Traces**
- **Hybride Ansätze**
- **Exploration vs Exploitation**

Diskretisierung von Zustands- und Aktionsräumen

Konferenzen, Symposia, Workshops
- *National conference on Artificial Intelligence (24th in 2010)*
- *Florida Artificial Intelligence Research Society (23th in 2010)*

Gliederung

Praktische Herausforderungen

Beschleunigung der Lernverfahren

- **Vorwissen über die Systemumgebung**

  vorprogrammierte, menschliche Vorkenntnisse
  geringerer Grad der Autonomie

Beispiele aus [Kaelbling u.a.]:
Erwartung bestimmter Funktionsverläufe (Approx.)
Passende, manuelle Diskretisierung des Zustandsraums
Vorkenntnisse über den Zustandsraum
Erfahrungen aus Statistiken

Gliederung

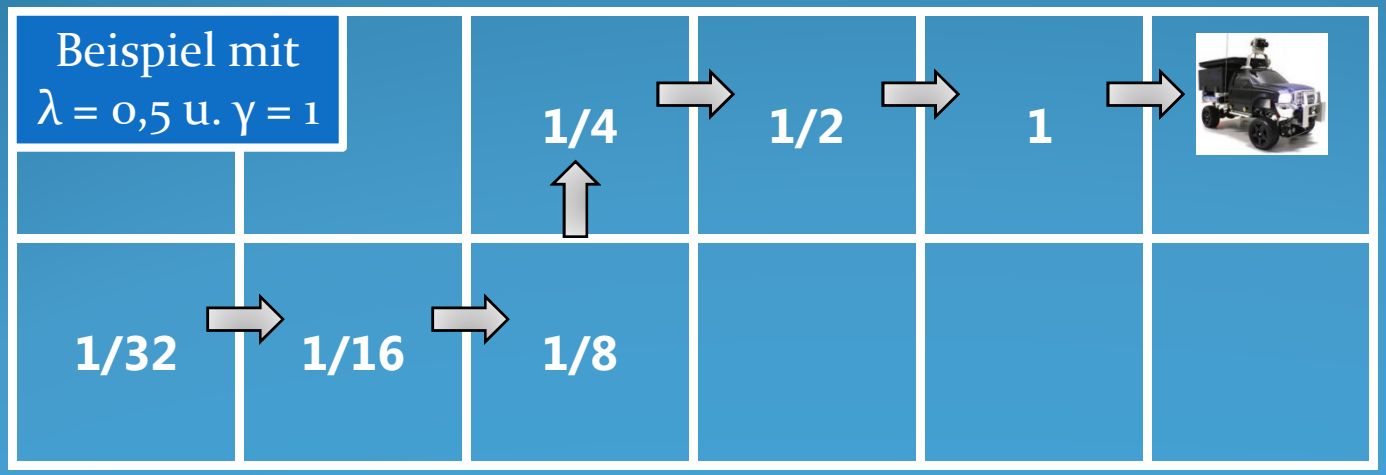
Praktische Herausforderungen

Beschleunigung der Lernverfahren

- **Eligibility Traces (*e-traces*)**

Erweiterung für TD Verfahren
Schnellere Erkundung des Zustandsraums durch Berücksichtigung von „*Delayed Rewards*“ und Einbeziehung der zuletzt besuchten Zustände

$$e_{t+1}(s) = \begin{cases} \lambda \cdot \gamma \cdot e_t(s), & falls\ s \neq s_t \\ 1, & sonst \end{cases}$$



Gliederung

Praktische Herausforderungen

Beschleunigung der Lernverfahren

- **Hybride Ansätze**

Häufig zur Gewinnung von Vorwissen
Hart kodierte, menschliche Vorkenntnisse (s.o.)

In [Thrasher u.a.]:
Kombination mit Supervised Learning
Gelerntes auf ähnliche Anwendungsfälle anwenden

Gliederung

Praktische Herausforderungen

Beschleunigung der Lernverfahren

- **Exploration vs Exploitation**

Methoden mit zufälliger Aktionsauswahl
- **ε-Greedy Suche**
- **Boltzmann Exploration**
- **Simulated Annealing**

Im Rahmen von SIRL* [Chen u. Dong]
- **Reinforcement Learning und Explorationen auf Grundlage von Quanten Charakteristiken**

*** SIRL: superposition-inspired reinforcement learning**

Gliederung

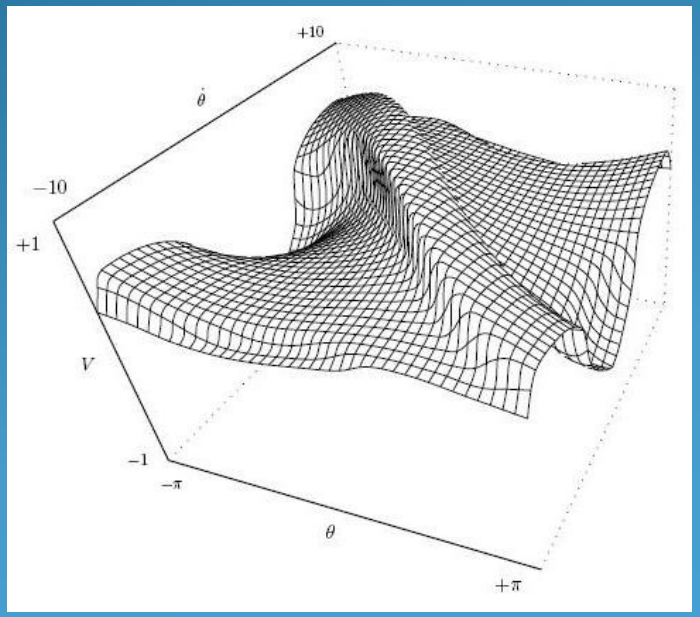
Praktische Herausforderungen

Diskretisierung von Zustands- und Aktionsräumen

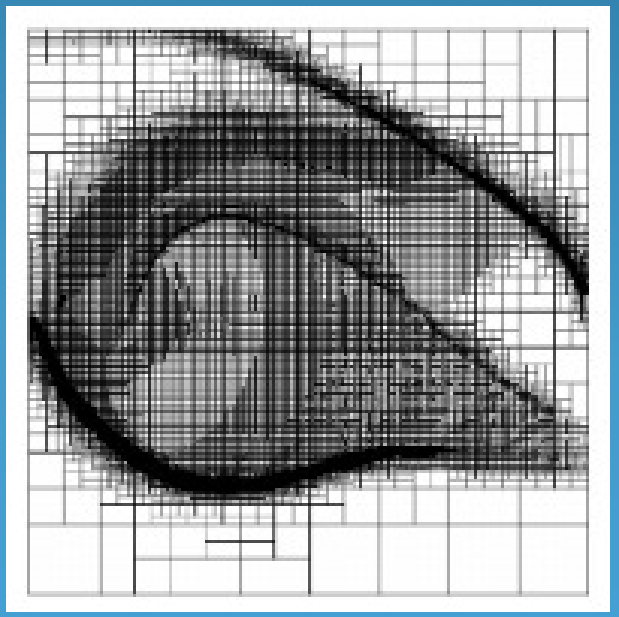
kontinuierliche Zustandsräume (z.B Winkel) oder Aktionsräume (z.B. Kräfte) praktisch unendlich groß

Nutzung der üblichen Methoden durch Diskretisierungen

Gitteransatz von [Coulom]



Gitteransatz fein/grob

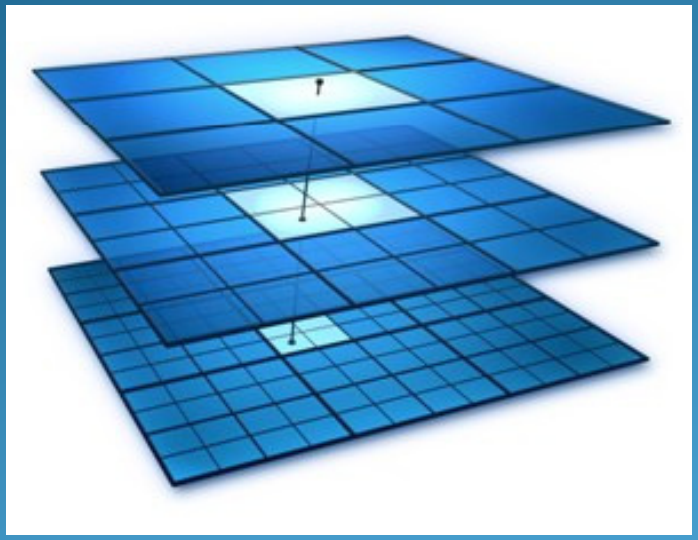


Gliederung

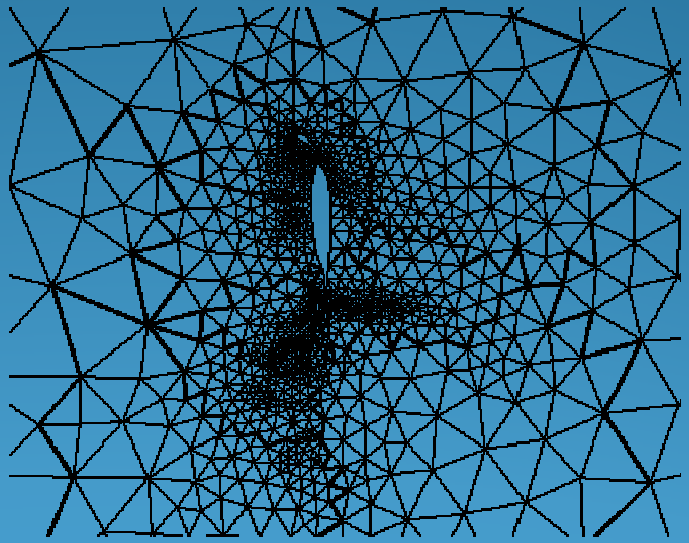
Praktische Herausforderungen

Diskretisierung von Zustands- und Aktionsräumen

Mehrgitterverfahren



Dreiecksverfahren



Gliederung

Zusammenfassung

unzählige, verschiedenartige Anwendungsfälle

in abgewandelten/erweiterten Formen

spezifische Finessen

Beschleunigung meist notwendig für praktikables Lernen

vielfältige Approximationsverfahren

grundlegende Vorgehensweise konkretisiert

ähnliche Anwendungsfälle als Anreize, Ideen

Gliederung

Bedeutung für eigene Arbeitsziele

keine aufwendigen Diskretisierungen erforderlich, aber

Approximationen z.B. für Sollgeschwindigkeiten möglich
- **Kubische Splines ?**
- **Radiale Basisfunktionen ?**

träges Beschleunigungsverhalten → Einfluss der Vorzustände
- **TD Verfahren bieten sich an**
- **Nutzung von e-traces**

Exploration zeitunkritisch und unabhängig von Exploitation

Vorwissen für schnellere Exploration nutzbar

irgendwann Selbstlokalisierung notwendig (Odometriefehler)

Reinforcement Learning Toolbox [Neumann]

Gliederung

Literaturverzeichnis

[Chen und Dong]
CHEN, Chun-Lin ; DONG, Dao-Yi: Superposition-Inspired Reinforcement Learning and Quantum Reinforcement Learning. In: *Reinforcement Learning: Theory and Applications*. Wien, Österreich : I-Tech Education and Publishing, 2008, S. 59–84. – ISBN 978-3-902613-14-1

[Coulom]
COULOM, M. R.: *Apprentissage par renforcement utilisant des réseaux de neurones, avec des applications au contrôle moteur*, Nationales Institut für Polytechnik Grenoble, Doktorarbeit, 2002

[Crites u. Barto]
CRITES, R.H.; BARTO, A.G.; *Improving elevator performance using reinforcement learning*. In Touretzky, D., Mozer, M., Hasselmo, M. (Eds.); Neural Information Processing Systems 8

[Kaelbling u.a.]
KAELBLING, Leslie P.; LITTMAN, Michael L.; MOORE, Andrew W.: Reinforcement Learning: A Survey. In: *Journal of Artificial Intelligence Research (JAIR)*. 4 (1996), S. 237-285

[Mahadevan]
MAHADEVAN, S.; CONNELL, J.; Automatic Programming of behavior-based robots using reinforcement learning. In: *Proceedings of Ninth National conference on Artificial Intelligence*, Anaheim, CA

Literaturverzeichnis

[Neumann]
NEUMANN, Gerhard: *The Reinforcement Learning Toolbox, Reinforcement Learning for Optimal Control Tasks*, University of Technology Graz, Diplomarbeit, 2005

[Sutton und Barto]
SUTTON, Richard S. ; BARTO, Andrew G.: *Reinforcement Learning - An Introduction*. Cambridge : MIT Press, 1998. – ISBN 978-0262193986

[Tesauro]
TESAURO, Gerald; Practical issues in temporal difference learning. *Machine Learning, 8,* 1992, S. 257-277
TESAURO, Gerald; Temporal difference learning and TD-Gammon. *Communications of the ACM, 38*(3), 1995, S. 58-67

[Thrasher u.a.]
THRASHER, Adam; ANDREWS, Brian; WANG, Feng: Control of FES using Reinforcement Learning: Accelerating the learning rate. In: *Proceedings – 19th International Conference – IEEE/EMBS*. Chicago, IL., USA : IEEE Computer Society, 1997, S. 1774-1776. – ISBN 0-7803-4262-3

[Williams]
WILLIAMS, R.J.; Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Machine Learning, 8,* 1992, S. 229-256

Vielen Dank für die Aufmerksamkeit!

Fragen?