**Universität Hamburg**
Department Informatik, AB TAMS
Norman Hendrich, Jianwei Zhang

Sekretariat: T. Tetsis
Raum F-311
Tel 42883-2430, Fax 42883-2397

# Übungen zum Modul WPM6: Algorithmisches Lernen

## SS 2010 Blatt 9

**Ausgabe:** 14 .07.2010, **Besprechung:** 14.07.2010

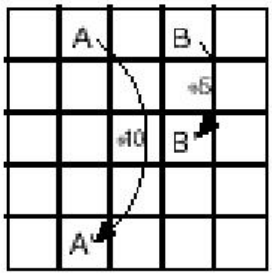
### Folgende Aufgaben sind freiwillig zur Übung gedacht

**Aufgabe 9.1 [1 Punkte] Grid World Programmierung:** Implementieren Sie mit einem Werkzeug Ihrer Wahl, **eine** der folgenden Aufgaben mit Hilfe der DP Algorithmen. $\gamma$ soll frei wählbar sein. Alle Zustände werden mit 0 initialisiert. Implementieren sie zunächsts die Policy-Evaluation für die Zufalls-Policy. Definieren sie ein $\Delta$, bei dessen Unterschreitung die Evaluierung terminiert. Führen Sie eine Policy-Verbesserung durch und erweitern Sie Ihr Programm um eine Policy Iteration.

Empfohlene Vorgehensweise für alle, die einen Einstieg suchen:

- Laden sie die Datei gridworld.c von der Seite der Übungen herunter
- Kompilieraufruf:
  gcc gridworld2.c -o gridworld2
- dieses Programm führt eine Policy Evaluation für das Beispiel Gridworld II durch

**9.1.1 [1]:** Das Grid World Beispiel I aus der Vorlesung. (RL1 - S. 58)

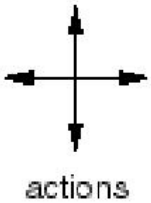




| | | | | |
|---|---|---|---|---|
| 3.3 | 8.8 | 4.4 | 5.3 | 1.5 |
| 1.5 | 3.0 | 2.3 | 1.9 | 0.5 |
| 0.1 | 0.7 | 0.7 | 0.4 | -0.4 |
| -1.0 | -0.4 | -0.4 | -0.6 | -1.2 |
| -1.9 | -1.3 | -1.2 | -1.4 | -2.0 |

- Eine kontinuierliche Aufgabe
- Aktionen, die den Agenten aus dem $Grid$ nehmen würden, lassen den Zustand unverändert und geben einen Reward von -1 (Ausnahme: die beiden Sprünge, die stets mit 5 und 10 belohnt werden)
- Der $Reward$ ist 0, außer wenn der Sprung A oder B ausgeführt wird

**9.1.2 [1]:** Das Grid World Beispiel II aus der Vorlesung. (RL2 - S. 9)




| | | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | |

$r = -1$
on all transitions

- Eine episodische Aufgabe
- Nicht-terminale Zustände :$1, 2, \ldots, 14$;
- Ein terminaler Zustand (zweimal als schattiertes Quadrat dargestellt)
- Aktionen, die den Agenten aus dem $Grid$ nehmen würden, lassen den Zustand unverändert

- Der $Reward$ ist - 1 bis der terminale Zustand erreicht ist

**9.1.3 [1]:** Ein Labyrinth Ihrer Wahl. Zeichnen Sie ein 3 x 3, 4 x 4 oder n x m Labyrinth mit Wänden, Sackgassen etc.

- Eine episodische Aufgabe
- Eine Start- und ein Zielzustand
- Gegen Wände laufen läßt den Zustand unverändert.
- Der $Reward$ ist -1 bis das Labyrinth gelöst ist