

# Reinforcement-Lernen

Wintersemester 09/10

Aufgabenzettel 3

**Abgabe:** 3.12.2010 **vor** der Vorlesung.

Abgabe von Programmieraufgaben per Email an `asja.fischer@ini.rub.de` .

## **Aufgabe 6** (100%). Q-learning und SARSA

Gegeben ist ein mit Fallen versehenes Labyrinth (`DangerousMaze.h`). Finde für diesen MDP eine optimale Policy

1. mit Hilfe von SARSA, und
2. mit Hilfe von Q-learning.
3. Erzeuge Lernkurven für Q-learning und SARSA und vergleiche die beiden Methoden. Die nach einer Anzahl von Episoden gelernte Q-Funktion kann jeweils auf zwei verschiedene Weisen ausgewertet werden. Daraus resultieren vier mögliche Vergleiche zwischen den beiden Algorithmen. Welcher Vergleich ist fair?

Verwenden zum Lösen der Aufgabe die Dateien `RLTask.h`, `DangerousMaze.h`, `dangerMaze.cpp` und `Policy.h`.

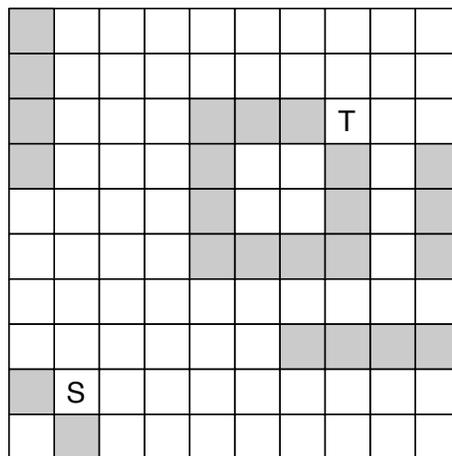


Abbildung 1: Grundriss des Labyrinths. Weiße Felder sind betretbar, graue Felder stellen Fallen dar. Der Startzustand ist mit **S** markiert und der Zielzustand mit **T**. Der Reward fuer Erreichen des Zielzustands beträgt 10, fuer Betreten einer Falle  $-50$  und sonst  $-1$ .