Reinforcement-Lernen

Wintersemester 10/11

Aufgabenzettel 2

Abgabe: 12.11.2010 **vor** der Vorlesung.
Abgabe von Programmieraufgaben per Email an `asja.fischer@ini.rub.de` .

Aufgabe 3 (20%)**.** Policy Improvement Theorem

In der Vorlesung wurde das Policy Improvement Theorem für deterministische Handlungsstrategien bewiesen. Beweise das Policy Improvement Theorem für stochastische Handlungsstrategien.

Aufgabe 4 (20%)**.** Gierige Handlungsstrategie (ϵ-greedy policy)

Wie bestimmt man eine gierige Handlungsstrategie basierend auf
a) der Zustandswertefunktion V?
b) der Zustandsaktionswertefunktion Q?
Betrachte auch den Fall, dass mehrere gierige Aktionen auftreten können.

Aufgabe 5 (60%)**.** Programmieraufgabe: on-policy Monte Carlo Control

Gegeben sei ein zweidimensionales Labyrinth (in der Datei `Maze.h`). Das Labyrinth ist in Felder eingeteilt. Die Position des Agenten (die Nummer des Feldes auf dem er gerade steht) beschreibt seinen Zustand. Auf jedem Feld stehen ihm die Aktionen *gehe nach Norden*, *gehe nach Osten*, *gehe nach Süden* und *gehe nach Westen* zur Verfügung. Der Agent erhält im terminalen Zustand den Reward $r = 10$ und in jedem nicht-terminalen Zustand einen Reward von $r = 0$, der "discount"-Faktor sei $\gamma = 0.9$.

Berechne die optimale Zustandsaktionswertefunktion Q mit *first-vistit on-policy Monte Carlo Control*. Verwende dazu die Dateien `onPolicyMonteCarlo.cpp`, `Maze.h` und `RLTask.h` und die Programmbibliothek **Shark**.

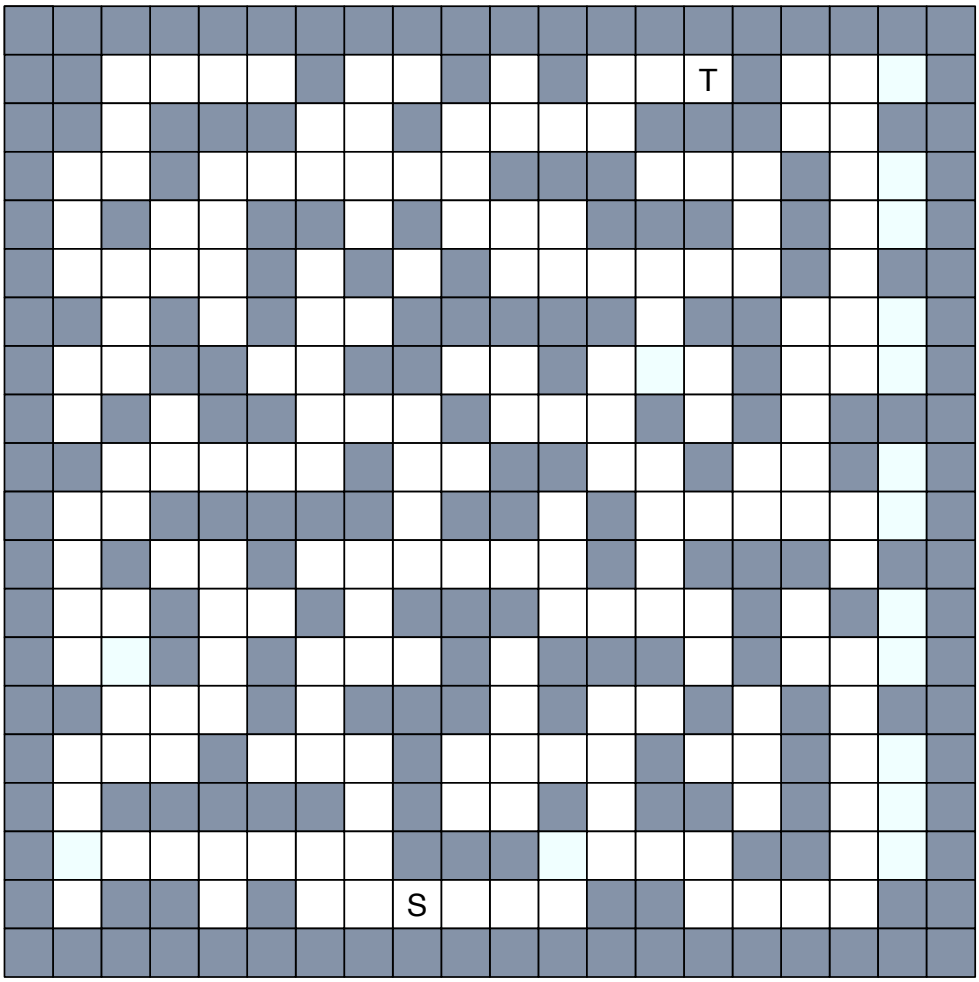

Abbildung 1: Grundriss des Labyrinths. Weiße Felder sind betretbar, graue Felder stellen Wände dar. Der Startzustand ist mit `S` markiert und der Zielzustand mit `T`.