

# Maschinelles Lernen II

## 1. Übung

Prof. Tobias Scheffer  
Dr. Niels Landwehr  
Christoph Sawade  
Uwe Dick  
Peter Haider

Sommer 2011

Ausgabe am: 28.04.11

### Aufgabe 1

*Recommendation Engine in RL-Glue implementieren*

Implementieren Sie einen Reinforcement-Learning-Agenten. Für die Implementierung wird das RL-Glue-Framework ([http://glue.rl-community.org/wiki/Main\\_Page](http://glue.rl-community.org/wiki/Main_Page)) verwendet. RL-Glue bietet die Möglichkeit, RL-Agenten in einer RL-Umgebung zu trainieren und zu evaluieren. Ihre Aufgabe wird es sein, einen Agenten zu programmieren, der in der Lage ist, eine Policy zu lernen, die besser ist als eine vorgegebene zufällige Policy.

Das Anwendungsgebiet ist das Anzeigen von personalisierten Bannern eines Online-Shops. Der Agent repräsentiert eine Recommendation Engine, die entscheidet, welche Banner des Online-Shops einem Nutzer angezeigt werden sollen. Jeder Banner entspricht einer Seite und einem Artikel des Online-Shops. Das Ziel des Agenten ist es, eine Policy zu lernen, die die Anzahl der Verkäufe maximiert.

Dem Agenten wird also von der Umgebung eine Sequenz von Beobachtungen präsentiert, d.h. besuchte Webseiten eines Nutzers, und nach jeder Beobachtung muss der Agent eine Aktion ausführen, d.h. einen Banner auswählen. Diese Problemstellung lässt sich als RL-Problem beschreiben, da davon ausgegangen werden kann, dass das temporal credit assignment-Problem auftritt. Denn auch ein nicht sofort angeklickter Banner kann den Nutzer dazu bewegen, sich den entsprechenden, oder einen ähnlichen, Artikel später anzugucken und zu kaufen.

Der Agent beobachtet die Seite, auf der der Banner angezeigt werden soll und die jeweilige Aktion, die auf diese Seite geführt hat. Beobachtungen werden repräsentiert als ein 2-Array von Integers. Der erste Integer ist der Index der Seite, der zweite beschreibt die Aktion des Users:

- 0 Abbruch. Der User verlässt den Shop.
- 1. Kauf. Der User kauft den Artikel.
- 2. Bannerklick. Der User hat auf den angezeigten Banner geklickt

3. Normaler Klick. Der User hat durch normale Navigation die Seite gewechselt.

Die Aktion des Agenten wird durch einen Integer beschrieben, der den Index der Seite repräsentiert, für die ein Banner angezeigt werden soll.

Es ist ein Beispielagent "Recom\_Agent.java" gegeben, der Banner zufällig auswählt. Ziel für die Evaluierung in der nächsten Woche wird es sein, einen Agenten zu implementieren, der eine Policy lernt, die höheren durchschnittlichen Reward erzielt als der Zufallsagent. Daten werden dem Agenten in Episoden gezeigt. Jede Episode entspricht einem Kunden des Onlineshops. Ein Experiment entsprechend dem RL-Glue-Format ist ebenfalls implementiert worden. "Recom\_Experiment.java" kann zum Testen beliebig verändert werden.

In der Übungsstunde am 05.05.11 wird der Lernalgorithmus des Agenten evaluiert. Dazu wird ein Experiment verwendet, das 50000 Episoden erzeugt. Bewertet wird der durchschnittliche Reward auf diesen 50000 Episoden. Es wird also keine Policy bewertet, sondern die Policy muss auf den Beispielepisoden gelernt werden.

Bei Fragen bitte direkt an Uwe Dick in Raum 0.19 wenden. (uwedick@cs.uni-potsdam.de)