



**School of  
Engineering**

InIT Institut für angewandte  
Informationstechnologie

## **Projektarbeit (Informatik)**

# Reinforcement Learning mit einem Multi-Agenten System für die Planung von Zügen

---

**Autoren**

---

Dano Roost  
Ralph Meier

---

**Hauptbetreuung**

---

Andreas Weiler

---

**Nebenbetreuung**

---

Thilo Stadelmann

---

**Datum**

---

18.09.2019

# Zusammenfassung

Zusammenfassung in Deutsch

# Abstract

Abstract in English

## **(Deutschsprachiges Management Summary)**

## **(Englischsprachiges Management Summary)**

# **Vorwort**

Stellt den persönlichen Bezug zur Arbeit dar und spricht Dank aus.

## Erklärung betreffend das selbständige Verfassen einer Projektarbeit an der School of Engineering

Mit der Abgabe dieser Projektarbeit versichert der/die Studierende, dass er/sie die Arbeit selbständig und ohne fremde Hilfe verfasst hat. (Bei Gruppenarbeiten gelten die Leistungen der übrigen Gruppenmitglieder nicht als fremde Hilfe.)

Der/die unterzeichnende Studierende erklärt, dass alle zitierten Quellen (auch Internetseiten) im Text oder Anhang korrekt nachgewiesen sind, d.h. dass die Projektarbeit keine Plagiate enthält, also keine Teile, die teilweise oder vollständig aus einem fremden Text oder einer fremden Arbeit unter Vorgabe der eigenen Urheberschaft bzw. ohne Quellenangabe übernommen worden sind.

Bei Verfehlungen aller Art treten die Paragraphen 39 und 40 (Unredlichkeit und Verfahren bei Unredlichkeit) der ZHAW Prüfungsordnung sowie die Bestimmungen der Disziplinar massnahmen der Hochschulordnung in Kraft.

Ort, Datum:

Unterschriften:

.....

.....

.....

.....

Das Original dieses Formulars ist bei der ZHAW-Version aller abgegebenen Projektarbeiten zu Beginn der Dokumentation nach dem Abstract bzw. dem Management Summary mit Original-Unterschriften und -Datum (keine Kopie) einzufügen.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>8</b>
1.1. Ausgangslage . . . . .	8
1.2. Zielsetzung / Aufgabenstellung / Anforderungen . . . . .	8
<b>2. Technical and mathematical foundation</b>	<b>9</b>
2.1. Reinforcement learning . . . . .	9
<b>3. Vorgehen / Methoden</b>	<b>10</b>
3.1. (Verwendete Software) . . . . .	10
<b>4. Resultate</b>	<b>11</b>
<b>5. Diskussion und Ausblick</b>	<b>12</b>
<b>6. Verzeichnisse</b>	<b>13</b>
Literaturverzeichnis . . . . .	13
(Abbildungsverzeichnis) . . . . .	15
(Tabellenverzeichnis) . . . . .	16
(Abkürzungsverzeichnis) . . . . .	17
(Listingverzeichnis) . . . . .	I
<b>A. Anhang</b>	<b>II</b>
A.1. Projektmanagement . . . . .	II
A.2. Weiteres . . . . .	II



# 1. Einleitung

## 1.1. Ausgangslage

- Nennt bestehende Arbeiten/Literatur zum Thema -> Literaturrecherche
- Stand der Technik: Bisherige Lösungen des Problems und deren Grenzen
- (Nennt kurz den Industriepartner und/oder weitere Kooperationspartner und dessen/deren Interesse am Thema Fragestellung)

## 1.2. Zielsetzung / Aufgabenstellung / Anforderungen

- Formuliert das Ziel der Arbeit
- Verweist auf die offizielle Aufgabenstellung des/der Dozierenden im Anhang
- (Pflichtenheft, Spezifikation)
- (Spezifiziert die Anforderungen an das Resultat der Arbeit)
- (Übersicht über die Arbeit: stellt die folgenden Teile der Arbeit kurz vor)
- (Angaben zum Zielpublikum: nennt das für die Arbeit vorausgesetzte Wissen)
- (Terminologie: Definiert die in der Arbeit verwendeten Begriffe)

## 2. Technical and mathematical foundation

### 2.1. Reinforcement learning

#### Basic definitions

In recent years, major progress in reinforcement learning has been achieved. In reinforcement learning, an agent  $\mathcal{A}$  learns to perform a task by interacting with an environment  $\mathcal{E}$ . On every discrete timestep  $t$  the agent needs to take an action  $u$ . This action  $u$  is based on the current observation  $s$ . If the agent does well, it receives positive reward from the environment, if it does something bad, there is no or negative reward. The goal of the agent  $\mathcal{A}$  is now to maximize the expected future reward  $\mathbb{E}[\mathcal{R}_{t+1} + \mathcal{R}_{t+1} + \mathcal{R}_{t+1} + \dots | s_t]$  given the current observation  $s$ .

The current observation  $s_t$ , also known as the current state is used to determine which action  $u$  to take next. An agent can observe its environment either fully or partially.

#### Value based versus policy based methods

Reinforcement learning methods are categorized into value-based methods and policy-based methods. Those variants differ on how they select an action  $u$  from a state  $s$ . Value-based reinforcement learning has its origins in dynamic programming. Through repeated rollouts of the environment, a value function  $\mathcal{V}(s)$  is acquired.  $\mathcal{V}(s)$  aims to estimate the future expected reward for any given state  $s$  as precisely as possible. This estimation  $\mathcal{V}(s)$  is achieved by either a lookup table for all possible states or a function approximator. In this work, we solely focus on the case that  $\mathcal{V}(s)$  is implemented in form of a neural network as function approximator. Using this approximation  $\mathcal{V}(s)$  we can now select the action  $u$  that takes the agent into the next state  $s_{t+1}$  with the highest expected reward

The second category of reinforcement learning algorithms are the so called policy based methods. These methods aim to acquire a stochastic policy  $\pi$  that maximizes the expected reward  $\mathcal{R}$  by taking actions with certain probabilities. Taking actions based on probabilities solves an important issue of value based methods, which is, that by taking greedy actions with respect to state  $s$ , the agent might not explore the whole state space and misses out on better ways to solve the environment (source!!).

Mnih et al, DQN Atari <https://www.cs.toronto.edu/~vmnih/docs/dqn.pdf>

Wu et al, A3C <https://arxiv.org/abs/1602.01783>

Overview over MARL, Hernandez-Leal et al <https://arxiv.org/pdf/1810.05587.pdf>

A3C in a multi agent environment, <https://arxiv.org/pdf/1903.01365.pdf>

## 3. Vorgehen / Methoden

- (Beschreibt die Grundüberlegungen der realisierten Lösung (Konstruktion/Entwurf) und die Realisierung als Simulation, als Prototyp oder als Software-Komponente)
- (Definiert Messgrößen, beschreibt Mess- oder Versuchsaufbau, beschreibt und dokumentiert Durchführung der Messungen/Versuche)
- (Experimente)
- (Lösungsweg)
- (Modell)
- (Tests und Validierung)
- (Theoretische Herleitung der Lösung)

### 3.1. (Verwendete Software)

Für die vorliegende Arbeit wurden die unten aufgeführten Programme eingesetzt.

#### Arbeitsumgebung

- Microsoft Windows 8 developer preview

#### Virtual Machine

- Oracle VM VirtualBox, Version 3.2.10

#### CAD Catia

- CATIA, Version 5.19 (in VirtualBox)

#### Dokumentation

- proTeXt mit TexMakerX 2.1 (SVN 1774), [latex-project.org](http://latex-project.org)
- Microsoft Visio 2007
- Adobe Acrobat 8 Professional 8.1.6

## 4. Resultate

- (Zusammenfassung der Resultate)

## 5. Diskussion und Ausblick

- Bespricht die erzielten Ergebnisse bezüglich ihrer Erwartbarkeit, Aussagekraft und Relevanz
- Interpretation und Validierung der Resultate
- Rückblick auf Aufgabenstellung, erreicht bzw. nicht erreicht
- Legt dar, wie an die Resultate (konkret vom Industriepartner oder weiteren Forschungsarbeiten; allgemein) angeschlossen werden kann; legt dar, welche Chancen die Resultate bieten

## **6. Verzeichnisse**

## Literaturverzeichnis

# **Abbildungsverzeichnis**



## **Tabellenverzeichnis**

## (Glossar)

In diesem Abschnitt werden Abkürzungen und Begriffe kurz erklärt.

Abk	Abkürzung
XY	Ix Ypsilon
YZ	Ypsilon Zet

## Listings

# A. Anhang

## A.1. Projektmanagement

- Offizielle Aufgabenstellung, Projektauftrag
- (Zeitplan)
- (Besprechungsprotokolle oder Journals)

## A.2. Weiteres

- CD mit dem vollständigen Bericht als pdf-File inklusive Film- und Fotomaterial
- (Schaltpläne und Ablaufschemata)
- (Spezifikationen u. Datenblätter der verwendeten Messgeräte und/oder Komponenten)
- (Berechnungen, Messwerte, Simulationsresultate)
- (Stoffdaten)
- (Fehlerrechnungen mit Messunsicherheiten)
- (Grafische Darstellungen, Fotos)
- (Datenträger mit weiteren Daten (z.B. Software-Komponenten) inkl. Verzeichnis der auf diesem Datenträger abgelegten Dateien)
- (Softwarecode)