

# (Deep) Reinforcement Learning understanding Policy Gradient

Thilo Stegemann  
Hochschule für Technik und Wirtschaft  
Master Student der Angewandten Informatik  
12459 Berlin, Wilhelminenhofstraße 75A  
Email: t.stegemann@gmx.de

*Abstract*—The abstract goes here.

## I. EINFÜHRUNG

This demo file is intended to serve as a “starter file” for IEEE conference papers produced under L<sup>A</sup>T<sub>E</sub>X using IEEE-tran.cls version 1.8b and later. I wish you the best of success.

mds

August 26, 2015

### A. Subsection Heading Here

Subsection text here.

1) *Subsubsection Heading Here*: Subsubsection text here.

## II. ERGEBNIS

The conclusion goes here.

## III. GRUNDLAGEN

Für ein besseres Verständnis des Policy Gradient Verfahrens wird nachfolgend ein Überblick der wichtigsten Grundlagen des Reinforcement Learnings (RL) gegeben. Das Problem des RL, das Agentenmodell, Maximierung der erwarteten Gewinne.

### A. Das Problem

Agent in einer Unbekannten Umgebung (Versuch und Irrtum)

### B. Komponenten eines RL Agenten

1) *Die Strategie (policy)*: Definiert das Verhalten des Agenten als Funktion. Maximierung der Summe der erwarteten Belohnungen entspricht der Findung einer optimalen Strategie

2) *Die Bewertungsfunktion (value function)*: Bewertet wie gut jeder Zustand und/oder jede Aktion ist.

3) *Das Modell*: Markov-Entscheidungsprozess Die vom Agenten genutzte Repräsentation der Umgebung.

## IV. VALUE-BASED REINFORCEMENT LEARNING

$$V_{\theta}(s) \approx V^{\pi}(s)$$

$$Q_{\theta}(s, a) \approx Q^{\pi}(s, a)$$

## V. POLICY-BASED REINFORCEMENT LEARNING

### VI. POLICY GRADIENT METHODEN

A. *Finite Difference Policy Gradient*

B. *Monte-Carlo Policy Gradient*

C. *Actor-Critic Policy Gradient*

## VII. THEORETISCHE ANWENDUNG NEURONALER NETZE

### ACKNOWLEDGMENT

The authors would like to thank...

### REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L<sup>A</sup>T<sub>E</sub>X*, 3rd ed. Harlow, England: Addison-Wesley, 1999.