Heidelberg University
Institute of Computer Science

Project report for the lecture Fundamentals of Machine Learning

Reinforcement Learning for Bomberman

`https://github.com/nilskre/bomberman_rl`

Team Member: Felix Hausberger, 3661293,
Applied Computer Science
eb260@stud.uni-heidelberg.de

Team Member: Nils Krehl, 3664130,
Applied Computer Science
pu268@stud.uni-heidelberg.de

Abstract

Plagiarism statement

Contents

List of Abbreviations

DQN	Deep-Q-Networks
MDP	Markov Decision Process

1 Introduction

2 Fundamentals and Related Work

Q-Learning is a known off-policy and model-free approach to train an agent based on temporal difference in an environment that can be modeled as a Markov Decision Process (MDP). An agent therefore does not necessarily use the policy it is trained for and does not know the transition probabilities and rewards in the MDP beforehand. Equation 1 shows the interative update formula for the Q-values that an online model uses to choose the right action.

$$Q_{k+1}(s,a) = (1-\alpha)Q_k(s,a) + \alpha(r + \gamma \max_{a'} Q_k(s',a')) \tag{1}$$

The problem with conventional Q-Learning is that in most of the cases the state dimension is far too high to explore and model the MDP entirely in foreseeable future. To deal with this problem the Q-values need to be approximated using a regression model. Deep neural networks have proven to be highly applicable for this task, which leads to the term of *Deep-Q-Learning* and respectively *Deep-Q-Networks* (DQN) for such network architectures. DQNs use the vectorized numerical state as its input and outputs the predicted Q-values. It learns through backpropagating the temporal difference error over each step. In the following papers regarding DQN architectures shall be introduced as well as papers dealing with the bomberman environment for reinforcement learning.

Paper [1] introduces ...

3 Approach

3.1 Reinforcement Learning Method and Regression Model

- Our model
- Our exploration method

3.2 Training process

- Experience buffer

4 Experimental results

5 Conclusion

A Appendix

A.1 Appendix A

References

[1] Hado van Hasselt, Arthur Guez, and David Silver. “Deep Reinforcement Learning with Double Q-learning”. In: *arXiv* (2015). URL: `https://arxiv.org/abs/1509.06461`.