Department of Physics and Astronomy
University of Heidelberg

Deep Q Learning Bomberman

Martina Sonja Zündel
Benjamin Raach
Henrik Ruh

WS 18/19

This report has been written by
Martina Sonja Zündel, Benjamin Raach, and Henrik Ruh
as a part of the Fundamentals of Machine Learning course
given by Prof. Köthe.

Abstract

Abstract.

1 Notes

tasks:

- write up
- think about how to perform a stagewise learning process
- reward system
- net architecture
- which loss criterion?
- which optimization algorithm?
- improve epsilon greedy implementation

  change of epsilon value as fct of number of rounds
- saving the model
- implement gpu ability
- implement terminal step learning
- perform learning
- code master script
- analyse agent decisions
- analyse learning progress
- cleaning up code

  commenting

  log messages

2 Results

2.1 learning scheme

- phase 1: no coins, no crates, no enemies
  - learn by reward to perform/ not perform a single action
  - learn to make a valid step (up, down, left, right) for the first action independent of the starting position
  - learn to make further non-invalid actions
- coins, no crates, no enemies
  - reward selecting coins
  - reward selecting coins in short time
- phase 2: coins, crates, no enemies
  - learn to make valid actions despite crates
  - reward destroying crates (without dying)
  - learn destroying crates to select coins
- phase 3: coins, crates, enemies
  - learn to avoid getting killed
  - learn to kill opponents
  - world dominance

2.2 Phase 1

effect of reward: empty map, no enemies, only coins: rewarding bombs -¿ after 100 training round, agent spams bombs penalizing bombs -¿ after 100 training round, agent goes always to the left penalizing invalid actions -¿ after 100 training round, agent goes always to the left -¿ after 1000 training round, agent goes always to UP