INF

FAKULTÄT FÜR
INFORMATIK

Emergence_RL

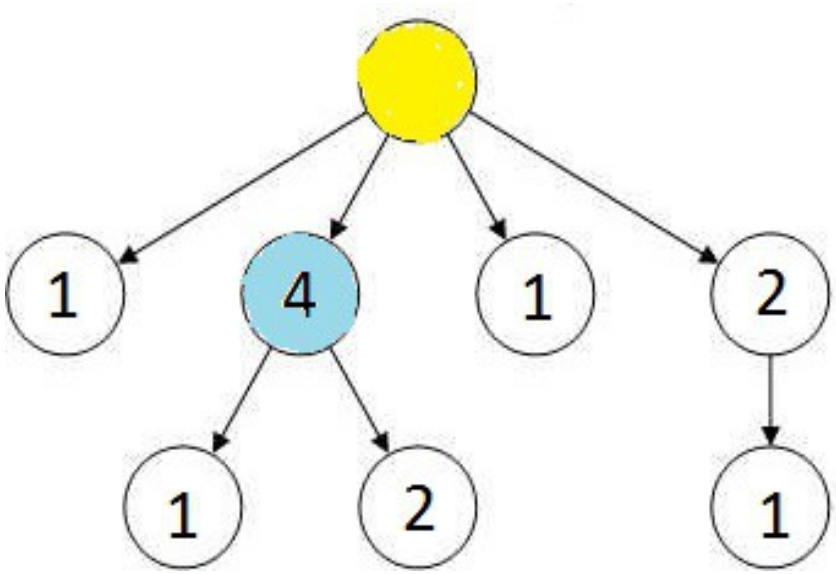
Julian Blank, Frederick Sander

Overview

- Actor and Backpropagation
- Heuristic tree policy
- Default policy
- Evolutionary algorithm

Actor and Backpropagation

- 1. Actor
  - Most visited Node



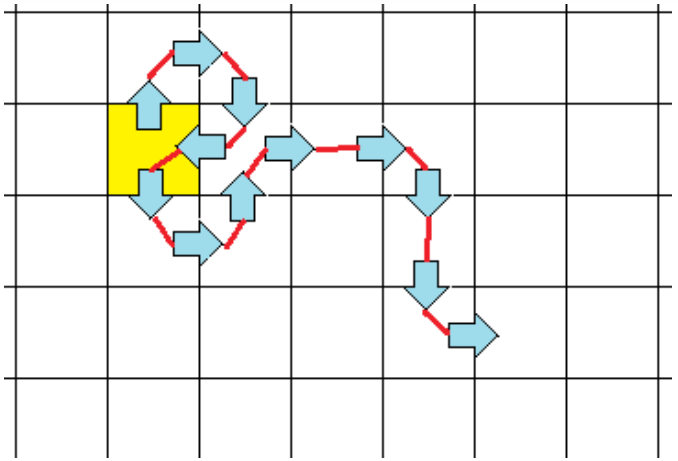
- 2. Backpropagation
  - Weighted reward:
    - Compute Actual-reward
    - Node.Reward = Node.Reward + Actual-reward
    - Node.Reward = Node.Reward * weight

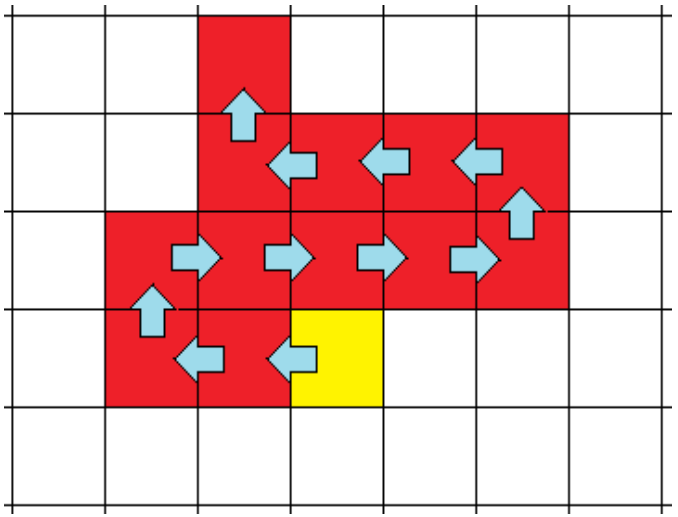
Heuristic tree policy

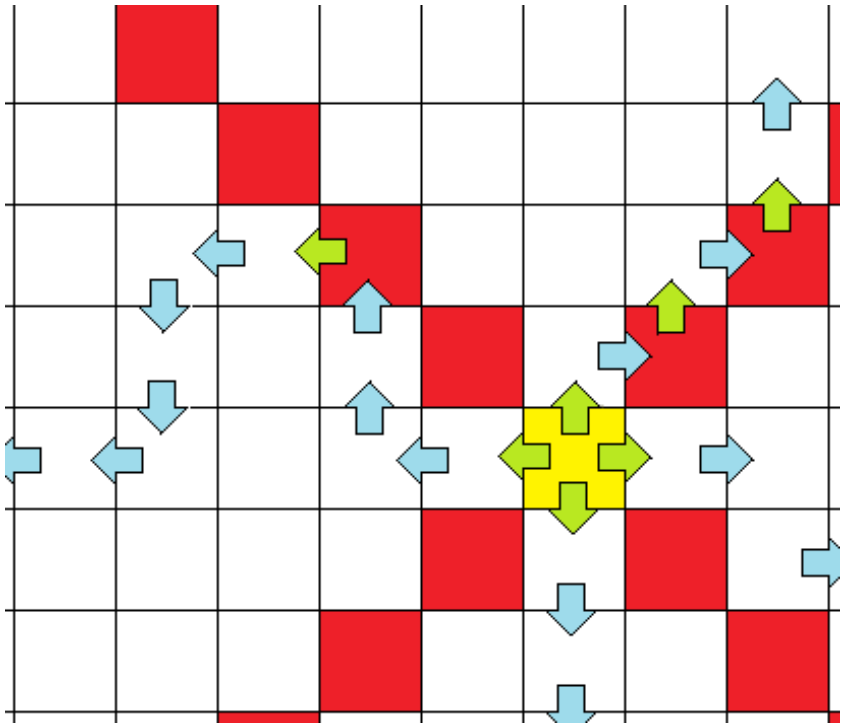
- Heuristic is used to compute the reward of a Node
  - Equation StateHeuristic
  - TargetHeuristic
- Heuristic tree policy
  - four weighted parameters:
    - Exploitation
    - Exploration
    - Heuristic value
    - History value
- Pessimistic Exploring:
  - Nodes with level ==1 (childs from the root) are tested more than once to improve safety

Default policy

- Different approaches:
  - Random walk
  - Self-avoiding walk
  - „FourRoomPolicy“







Evolutionary Algorithm

- Used an EA to find the best heuristic
  - Is generated online, while the game is running
  - some time is reserved to compute steps
    - in the constructor from the Agent and
    - every gametick
- After a defined number of timesteps, the actual heuristic is replaced by the new one

References

- http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6145622&tag=1
- http://mathworld.wolfram.com/Self-AvoidingWalk.html