

TUD Poker Challenge 2008

Reinforcement Learning with Imperfect Information

Outline

- Reinforcement Learning
- Perfect Information
- Imperfect Information
- Lagging Anchor Algorithm
 - Matrix Form
 - Extensive Form
- Poker Game
- Tools and Sources

Reinforcement Learning

- RL is sub-area of machine learning
- Basic reinforcement learning model consists of:
 - a set of environment states S
 - a set of actions A
 - a set of scalar "rewards" in R

Reinforcement Learning

- At each time t , the agent perceives its state $s_t \in S$ and the set of possible actions $A(s_t)$
- It chooses action $a \in A(s_t)$ and receives from the environment the new state s_{t+1} and a reward r_{t+1} .
- RL agent must develop a policy $\pi: S \rightarrow A$ which maximizes the quantity R for Markov Decision Processes (MDPs)

Perfect Information

- ❑ Chess and Backgammon are games with perfect information
- ❑ Time Difference (TD)-learning and Q-learning are used for games with perfect information

Perfect Information

- Main goal is finding the optimal policy in the policy space
- „Gradient descent“ as an optimization algorithm for finding a local minimum
- Temporal Difference- Learning algorithm can be constructed from the Bellman Equation through replacing expectations with estimates and then performing gradient descent

Imperfect Information

- ❑ Since poker is a card game, the current state of the game is hidden
- ❑ Poker is a game with imperfect information

Imperfect Information

- ❑ No exact calculation of the solution is possible
- ❑ Simple gradient search oscillates around the solution points
- ❑ Approximation technique is needed
- ❑ Lagging anchor algorithm is useful for the approximation

Lagging Anchor Algorithm

- ❑ Idea is to have an „anchor“ for each player which is lagging behind the current values of the parameter states
- ❑ Lagging anchor is dampening the oscillation of the simple gradient search
- ❑ Goal is to find the minmax solution point
- ❑ The algorithm can be implemented for games in matrix form and extensive form

Matrix Form

- ❑ Selten's anticipatory learning rule is used
- ❑ Algorithm produces approximate solutions to large games with non-linear and incomplete parameterization

Extensive Form

- ❑ The process of estimating the gradient is split into two
- ❑ First estimate gradient of expected payoff with respect to it's action probabilities
- ❑ Then calculate gradient of the agents action probabilities with respect to it's parameters

Poker game

- Set of possible actions A consists of:
 - Fold
 - Call
 - Raise

Poker game - Model

- ❑ Player and opponent are modeled through NN
- ❑ Evaluator is modeled through NN
- ❑ Game is modeled through NN
- ❑ Result is evaluator is used to train the player against the opponent

Tools and Sources

- ❑ <http://www.cse.unsw.edu.au/~cs9417ml/RL1/sourcecode.html>
- ❑ <http://www.cs.cmu.edu/~awm/rlsim/>
- ❑ <http://rlai.cs.ualberta.ca/RLAI/rlai.html>
- ❑ <http://www.sourceforge.net/projects/piql>

Questions

Thanks for your attention!