Research Proposal: Option Transfer Learning for General Video Game Playing

Maarten de Waard
Student number: 5894883

May 14, 2015

Many AI solutions exist to solve specific tasks. Several real world problems can be modelled in a game form. Therefore, solving games can be viewed as a stepping stone to solving more complex real world problems. A generic algorithm that can solve many games, regardless of their inner dynamics, can be viewed as a good, multi-purpose algorithm. A set of challenging tasks is the set of Arcade Games, which range from long term planning tasks like *Boulder Dash*, a puzzle game in which gems have to be collected in the right order, to quick reaction tasks like *Space Invaders*, a game in which aliens have to be shot from the sky as quickly as possible, while evading their missiles.

A framework called *Video Game Description Language (VGDL)* enables easy modelling of arcade games. The framework provides an easy way to test a reinforcement learning algorithm's performance on all the games that can be implemented in the framework. GVG-AI[1] is a competition for solving a set of games that are defined within the VGDL framework. The competition sets boundaries to the amount of computation time that is allowed when starting a game and choosing an action.

The aim of this study is to make an algorithm that can easily solve all games that range from puzzle games like Boulder Dash, to action games like Space Invaders. The games will be modelled as *Markov Decision Processes (MDPs)*. Traditionally, planning in MDPs tries to maximize cumulative reward by choosing optimal actions in states. This thesis will introduce *options*[1] into the GVG-AI competition, which means that planning will be done over sequences of actions, instead of a single action at a time. Using options can be seen as a more high-level approach to planning. Note that this eliminates the Markov-property from the MDP, resulting in what is called a *Semi-Markov Decision Process (SMDP)*. The hypothesis is that by adding higher level information a reinforcement learning algorithm can more effectively search the search space of the MDP, resulting in a higher cumulative reward.

Several options will be implemented, after which they are embedded in Reinforcement Learning algorithms *Q-learning* and *Monte Carlo Tree Search (MCTS)* in order to test the hypothesis. These will then be compared to the original Q-learning and MCTS algorithms without options, using the game score as a measure of performance. Good working solutions will be entered into the GVG-AI competition, comparing it to several other algorithms.

Planning

This thesis has started around February. Due to the Robocup Iran Open, a small delay has introduced itself. Therefore this is my rough planning:

- February - April: Reading into the subject
- May - July: Programming, Testing hypothesis, starting writing
- August - October: Writing

References

[1] R. S. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1):181–211, 1999.

[1] `gvgai.net`