COMP3130 – Group Project in Computer Science
10×10 Othello Learning Agent

Andrew Haigh – u4667010;
Timothy Cosgrove – u4843619;
Joshua Nelson – u4850020

June 8, 2012

1. Abstract

An agent to play the board game *Othello* is created, with the ability to learn through reinforcement. The minimax algorithm is used for game playing, and a static evaluation function for the leaf nodes is learnt by self play. The agent learns the insignificance of the number of stones, and the significance of stone positioning. This agent is played against itself, and against other developed agents, and it's performance is analysed.

2. Problem overview

3. Solution overview

4. Optimisations

5. Static evaluation function

6. Learning

6.1. TD-λ

The TD-λ algorithm was used to learn the weighting of the static evaluation function's various features. The implementation of this algorithm is based on the one implemented by the Knightcap agent (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.140.2003) At the end of each game played, the agent adjusts it's weights according to the following formula

$$w := w + \alpha \sum_{t=1}^{N-1} \Delta\tilde{J}(x_t, w) \;\times\; \left[\sum_{j=t}^{N-1} \lambda^{j-t} d_t\right]$$

w	The vector of weights	x_t	The t^{th} board in the game
α	The learning rate	λ	The discount factor
N	The number of states in the game	d_t	$\tilde{J}(x_{t+1}, w) - \tilde{J}(x_t, w)$
$\Delta\tilde{J}$	The derivative of the $\tilde{J}$ function		

In the above formula, the $\tilde{J}$ function estimates the probability of winning from a given state, given a set of weights for features and our board. It approximates the J function,

$$J(x_t) = \begin{cases} 1, & \text{if } x_t \text{ is a winning state} \\ 0, & \text{if } x_t \text{ is a lost state} \end{cases}$$

For each game state x_t, we adjust the weights according to a factor d_t, which is the temporal difference. $d_t = \tilde{J}(x_{t+1}, w) - \tilde{J}(x_t, w)$, and the weight adjustment is scaled with this amount. The motivating observation is, for the true J function, $J(x_{t+1}) - J(x_t) = 0$, so d_t should be as close to zero as possible.

6.2. ELO arena

6.3. Comparison of TD-λ and ELO arena

7. Performance evaluation

8. Improvements