EPT (Early Playout Termination) in MCTS

or

Lessons learned applying EPT to Amazons, Breakthrough, and Havannah

Richard Lorentz

Department of Computer Science
California State University
Northridge (Los Angeles), California
U.S.A.
What is EPT?

- MCTS algorithm (diagram from Chaslot's thesis)
What is EPT?

- Terminate the simulation/playout/rollout early and evaluate
- Result of evaluation reports win or loss – boolean result
History


  **Amabot**

- Winands M. H.M., Björnsson, Y.: Evaluation Function Based Monte-Carlo *LOA*. In Proceedings of the Twelfth International Advances in Computer Games Conference (ACG’09), Pamplona, Spain (2009)


  **Breakbot**

- Havannah (in progress)

  **Havbot**
Amabot

- Started as a mini-max program 2001
  - Numerous 2\textsuperscript{nd} and 3\textsuperscript{rd} place finishes at the Olympiads
  - Always behind Johan de Koning's 8QP
- Developed MCTS Amabot in 2007
  - MCTS was in full swing
  - Evolved to EPT (as did Campya)
  - 5 consecutive Olympiad golds
  - Undefeated on Little Golem
Amazons – starting position
Amazons – after 2 moves
Amazons – near the end
Amazons – “final” position
Breakbot

- Started as an MCTS program in 2012
  - Fully expected EPT transition
  - Little Golem rating increased from 1800 to 2100
  - Outperforms mini-max based programs on LG
  - 3rd rated program on LG
Breakthrough – initial position
Breakthrough – before first capture
Breakthrough – White resigns
Havbot

- Started as an MCTS program in 2012
  - Motivated by the “Havannah Challenge”
- Mini-max programs doomed (e.g. PZN)
- Reasonably strong on sizes 4 – 7
- Reasonably miserable on sizes 8 – 10
  - Progress slow for large sizes
- EPT created in 2014
  - 60% win rate against MCTS (size 8)
  - 50% win rate against MCTS (size 6)
    - Easier to play well on small boards
  - Game play is more natural
Havana – early position
Havana – White is looking good
Havana – Black wins
Blending EPT with mini-max

- Switch to mini-max near the end of a game? MCTS can easily miss a good tactical move.
  - Havbot & Breakbot have MCTS solvers, so no need
  - Amabot
    - Does not have a solver
      - Game ends well before the last move
    - Switched to mini-max when “game progress” was sufficiently large
  - Results were unclear, but because of defective territory EPT (mainly MCTS) is to be preferred
- Mini-max evaluation – hard to deal with defects
- MCTS and EPT – handles them well
  - Random playouts sample the defective area
Progressive Widening or New Node Count Initialization

- Amabot
  - Progressive widening based on move evaluation
  - Forward pruning (unsafe) – top 750 moves
- Havbot & Breakbot
  - Count initialization
    - Based on move features rather than (slow) evaluation values
    - Havbot: joints, no peek, etc.
    - Breakbot: (safe) progress towards goal, capture
When to Terminate?

- Near the end of the game seemed reasonable
- Amabot, Havbot, & Breakbot all had almost identical results
- Breakbot:

<table>
<thead>
<tr>
<th>Termination</th>
<th>Winning Result (as Black)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33%</td>
</tr>
<tr>
<td>4</td>
<td>43%</td>
</tr>
<tr>
<td>6</td>
<td>27%</td>
</tr>
<tr>
<td>12</td>
<td>10%</td>
</tr>
</tbody>
</table>
When to Terminate?

- Why not EPT > 4?
  - Not surprising – Deep playout dilutes reliability of evaluation

- Why not EPT < 4?
  - Is surprising!
  - Compare moves: termination = 4 vs. termination = 1
    - Move selection differs 12 times per (55 move) game
    - Differs 7 times per game just because of MCTS
    - Therefore ~ 5 significant differences per game = 10% of the moves
When to Terminate?

- How do EPT 4 moves compare with EPT 1 moves?
  - Disadvantage:
    - 4 random moves lose information
    - But, less serious in stable positions
  - Advantage:
    - Evaluation more accurate later in the game
    - MCTS will expand around unstable positions

- Optimum point around 4 or 5
Improved Playouts

- Crucial in pure MCTS
- Little to no benefit in EPT beyond the obvious choices
  - Amabot
    - Not used at all
  - Havbot
    - Make winning moves
    - Encourage joints and moves near other pieces
    - Nothing else helps
  - Breakbot
    - Make winning moves and easily proved forced wins
    - Encourage recaptures, safe moves, and deep moves
    - Nothing else helps
Evaluation Parity Effect

- Evaluation can vary wildly depending on whose turn to move
  - Always evaluate with the same person to move
  - More important at faster time controls

- Amabot
  - A small but noticeable advantage
    - Mainly play real time
    - ~ 2% increased win rate
    - Evaluation has strong parity effect

- Havbot & Breakbot
  - Evaluation only shows some parity effect
  - Most games played with large time controls (turn based game server (Little Golem))
Tweaks

- MCTS gets stuck on a suboptimal move
  - “Satisfactory” move can overwhelm other, better moves
  - Good tactical moves get insufficient wins
  - Can happen even with very large time controls
- Tweak: Increase UCTK at the root
  - Statistically inconclusive
  - In practice, rescued some games
  - More likely to help than to do damage
  - Unclear how to generalize in situations with no UCTK, e.g., RAVE
Tweaks

- Put strategic ideas in MCTS selection phase
  - Breakbot
    - Captures (not recaptures) tend to be bad early in the game
    - Hard to capture idea in evaluation
    - Modify the selection phase of MCTS:
      
      \[
      \text{move\_val} = \frac{\text{wins} - \text{capture\_penalty}}{\text{visits}} + \text{uct}
      \]
  
  - Havbot
    - Discourage useless peeks – failed in self tests 😞

- Amabot
  - Not used
Tweaks

- Evaluation function speed (compare with playout speed)
  - Amabot – less accurate but faster evaluation outperforms slow but precise evaluation in fast games
    - Gaining extra knowledge is expensive

- Breakbot & Havbot – improving evaluation worth extra cost in speed
  - Especially true at slow time controls
  - Evaluations are rather primitive
What Next?

- Amabot – plays very well, so progress comes slowly
  - Improve evaluation?
  - Quite sophisticated: developed 2001 – 2006
    - Square control
    - Territory
    - Piece mobility
    - Balancing the 3 according to which is dominant and the stage of the game
What Next?

- Breakbot – plays well, progress still slow
  - Improve evaluation?
    - Rather simple.
      - Advancing is good
      - Exchanging can be bad (tempo loss)
    - White's first row is good
  - Improve strategy
    - Avoid holes, especially in the middle
    - Zugzwang ...
What Next?

- Havbot – plays lousy
  - Size 4 – 6 pretty good
  - Size 7, 8 mixed
  - Size 9, 10 miserable
- Solve stupid peek problem
- Solve trapped pieces problem
- How?
  - Evaluation – still quite simple
    - Estimated distance to win
  - Node initialization
    - Not likely – bad moves often enter late
  - MCTS selection
    - First attempt failed
Conclusions

- EPT can improve (mini-max) programs with strong evaluations
- EPT can improve (MCTS) programs by introducing a weak evaluation
- EPT can and should be the algorithm of choice in many cases