BomberMan Tournament

Mustafa Fuad Rifet Ibrahim

Overview

- Game + Tournament
- 3 Agents
Game + Tournament

- Environment (crates, walls, coins,...)
- Agents (move, drop bombs)
- score
Game + Tournament

- Multiple eps
- 400 steps
- 0.5s thinking time
- Slow thinking penalty
- No multiprocessing
- One core of i7-8700K, up to 8GB RAM
- Simple agents

http://randomhooahaas.flyingomelette.com/bomb/gba-st/img/art-multiplay.JPG
Lord_Voldemort

https://cdn.wallpapersafari.com/6/42/hSU3zx.jpg
$Q(s, a) = (1 - \eta)Q(s, a) + \eta \left[ r + \gamma \max_{a'} Q(s', a') \right]$
state = (Left, Up, Right, Down, Self, Self_Bomb)
Lord_Voldemort

\[\text{state} = (Left, Up, Right, Down, Self, Self\_Bomb)\]

Self:
• Empty
• Danger
• Bomb
Lord_Voldemort

\[\text{state} = (\text{Left}, \text{Up}, \text{Right}, \text{Down}, \text{Self}, \text{Self\_Bomb})\]

Self:
- Empty
- Danger
- Bomb

Self\_Bomb:
- True
- False
Lord_Voldemort

\[ \text{state} = (\text{Left, Up, Right, Down, Self, Self\_Bomb}) \]

Left, Up, Right, Down:
- Wall
- Enemy
- Crate
- Coins
- Bombs
- Danger
- Empty
- Priority

Self:
- Empty
- Danger
- Bomb

Self\_Bomb:
- True
- False
Lord_Voldemort
Lord_Voldemort

<table>
<thead>
<tr>
<th>MATCH</th>
<th>ROUNDS</th>
<th>KILLS</th>
<th>SUICIDES</th>
<th>COINS</th>
<th>SCORE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group E</td>
<td>01</td>
<td>25</td>
<td>13</td>
<td>2</td>
<td>132</td>
<td>197</td>
</tr>
<tr>
<td>Group E</td>
<td>02</td>
<td>25</td>
<td>24</td>
<td>3</td>
<td>129</td>
<td>249</td>
</tr>
<tr>
<td>Group E</td>
<td>03</td>
<td>25</td>
<td>4</td>
<td>1</td>
<td>134</td>
<td>154</td>
</tr>
<tr>
<td>Group E</td>
<td>04</td>
<td>25</td>
<td>27</td>
<td>0</td>
<td>221</td>
<td>356</td>
</tr>
<tr>
<td>Octofinal J</td>
<td>100</td>
<td>61</td>
<td>21</td>
<td>21</td>
<td>607</td>
<td>912</td>
</tr>
<tr>
<td>Quarterfinal R</td>
<td>100</td>
<td>35</td>
<td>22</td>
<td>22</td>
<td>297</td>
<td>472</td>
</tr>
<tr>
<td>Semifinal V</td>
<td>100</td>
<td>48</td>
<td>14</td>
<td>14</td>
<td>319</td>
<td>559</td>
</tr>
</tbody>
</table>

https://hci.iwr.uni-heidelberg.de/vislearn/HTML/teaching/courses/FML/bomberman_rlagent.php?id=Lord_Voldemort
\[ L = \frac{1}{2} \left[ r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]^2 \]
NOBEL

Fixed Q-Targets

\[ L = \frac{1}{2} \left[ r + \gamma \max_{a'} Q(s', a') - \underbrace{Q(s, a)}_{\text{target}} \right]^2 \]
Dueling DQN

\[ Q(s, a) = V(s) + [A(s, a) - \max_{a'} A(s, a')] \]

\[ Q(s, a) = V(s) + [A(s, a) - \frac{1}{|A|} \sum_{a'} A(s, a')] \]

Wang et al., Dueling Network Architectures for Deep Reinforcement Learning, 2016
Wang et al, Dueling Network Architectures for Deep Reinforcement Learning, 2016
(Prioritized) Experience Replay:
• Data efficiency
• Learning the whole game
• Valuable experiences
<table>
<thead>
<tr>
<th>MATCH</th>
<th>ROUNDS</th>
<th>KILLS</th>
<th>SUICIDES</th>
<th>COINS</th>
<th>SCORE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group E</td>
<td>01</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>Group E</td>
<td>02</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>46</td>
<td>51</td>
</tr>
<tr>
<td>Group E</td>
<td>03</td>
<td>10</td>
<td>5</td>
<td>5</td>
<td>33</td>
<td>54</td>
</tr>
<tr>
<td>Group E</td>
<td>04</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>Group E</td>
<td>05</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Group E</td>
<td>06</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Group E</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>54</td>
<td>59</td>
</tr>
<tr>
<td>Group E</td>
<td>12</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td>Group E</td>
<td>13</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td>Group E</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>69</td>
<td>69</td>
</tr>
<tr>
<td>Octofinal L</td>
<td>100</td>
<td>32</td>
<td>25</td>
<td>481</td>
<td>641</td>
<td>1</td>
</tr>
<tr>
<td>Quarterfinal R</td>
<td>100</td>
<td>41</td>
<td>38</td>
<td>255</td>
<td>360</td>
<td>3</td>
</tr>
</tbody>
</table>

LaranTu

- Make it simple --> minigame
- Heuristic to get there
\[ v_i + C \times \sqrt{\frac{\ln N}{n_i}} \]
LaranTu

Selection based on:
• Value
• Policy prediction
• # Visited
LaranTu

Conv Batchnorm ReLu → Conv Batchnorm ReLu → Conv Batchnorm ReLu

(2x) → ReLU

Conv Batchnorm ReLu → Conv Batchnorm ReLu → Conv Batchnorm ReLu

FC ReLU → FC ReLU → FC Tanh

Prob. for 6 actions

Value of game state
LaranTu

Silver et al. 2017, "Mastering the game of Go without Human Knowledge"
<table>
<thead>
<tr>
<th>MATCH</th>
<th>_rounds</th>
<th>KILLS</th>
<th>SUICIDES</th>
<th>COINS</th>
<th>SCORE</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group D</td>
<td>03</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Group D</td>
<td>05</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Group D</td>
<td>06</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>-2</td>
</tr>
<tr>
<td>Group D</td>
<td>08</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Group D</td>
<td>09</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Group D</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Group D</td>
<td>12</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Group D</td>
<td>13</td>
<td>10</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Group D</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Group D</td>
<td>15</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>-3</td>
</tr>
<tr>
<td>Octofinal J</td>
<td>100</td>
<td>1</td>
<td>25</td>
<td>48</td>
<td>-43</td>
<td>4</td>
</tr>
</tbody>
</table>

https://hci.iwr.uni-heidelberg.de/vislearnHTML/teaching/courses/FML/bomberman_rl/agent.php?id=LaranTu
Sources

• Report: LaranTu, NOBEL, Lord_Voldemort
• neuro.cs.ut.ee/demystifying-deep-reinforcement-learning/
• Wang et al, Dueling Network Architectures for Deep Reinforcement Learning, 2016
• Silver et al, A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, 2018