

Theory and Practice of Artificial Intelligence

Further Games

Daniel Polani

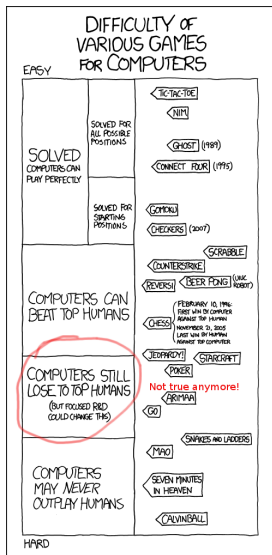
School of Computer Science
University of Hertfordshire

March 9, 2017

All rights reserved. Permission is granted to copy and distribute these slides in full or in part for purposes of research, education as well as private use, provided that author, affiliation and this notice is retained. Some external illustrations may be copyrighted and are included here under "fair use" for educational illustration only.

Use as part of home- and coursework is only allowed with express permission by the responsible tutor and, in this case, is to be appropriately referenced.

Obligatory XKCD



<https://xkcd.com/1002/> (CC BY-NC 2.5)

- one of the great breakthroughs in game AI
- based on exploration/exploitation tradeoffs regret (Auer 2003)
- generalized to trees (Kocsis and Szepesvári 2006)

Note: do not have the time for the full theory
just sketch the method

UCT Monte Carlo Tree Search II

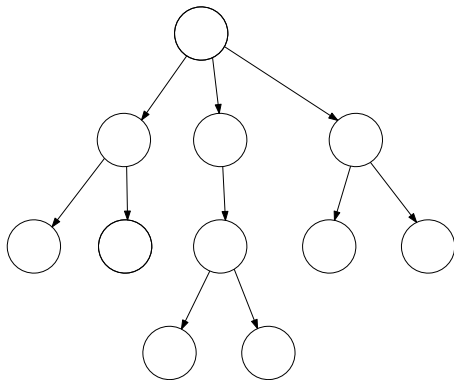
(Browne 2012; Browne et al. 2012; Bradberry 2015)

- Outset:**
- consider an **already expanded partial tree**
 - assume every node contains a
 - sum of rewards $\sum V_i$ hitherto collected from nodes beneath it
 - number of runs n that went through that node

for now, just a search, will generalize to games later

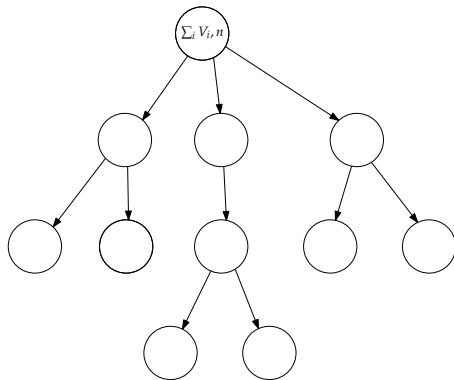
UCT Monte Carlo Tree Search III

(Browne 2012; Browne et al. 2012; Bradberry 2015)



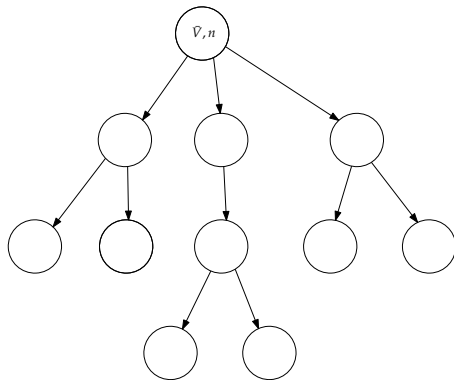
UCT Monte Carlo Tree Search IV

(Browne 2012; Browne et al. 2012; Bradberry 2015)



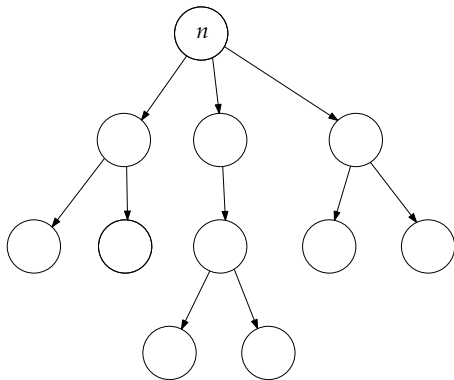
UCT Monte Carlo Tree Search V

(Browne 2012; Browne et al. 2012; Bradberry 2015)



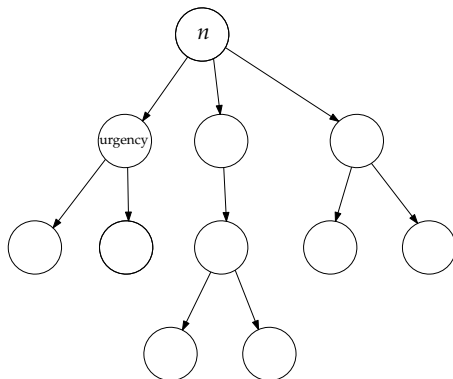
UCT Monte Carlo Tree Search VI

(Browne 2012; Browne et al. 2012; Bradberry 2015)



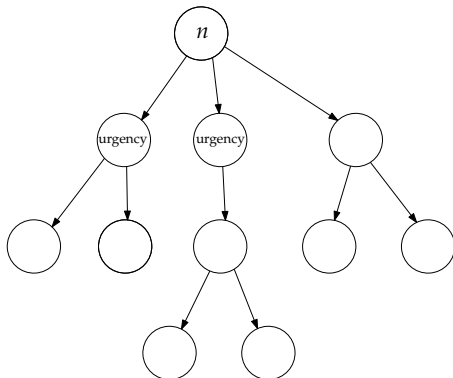
UCT Monte Carlo Tree Search VII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



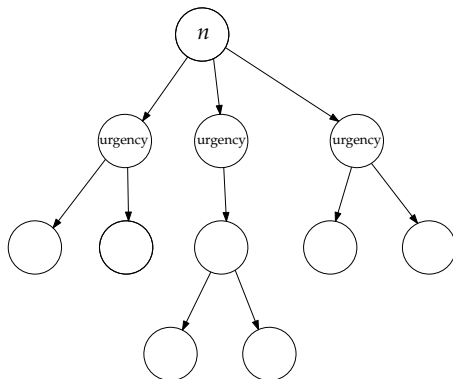
UCT Monte Carlo Tree Search VIII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



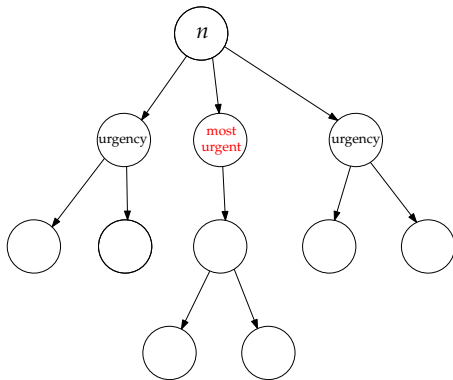
UCT Monte Carlo Tree Search IX

(Browne 2012; Browne et al. 2012; Bradberry 2015)



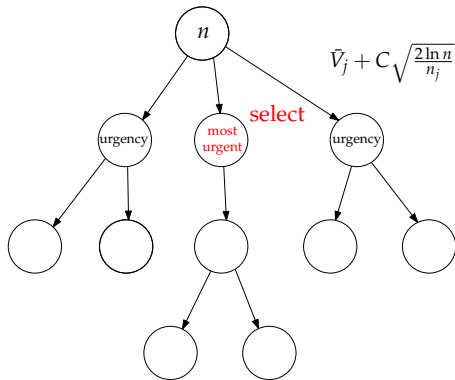
UCT Monte Carlo Tree Search X

(Browne 2012; Browne et al. 2012; Bradberry 2015)



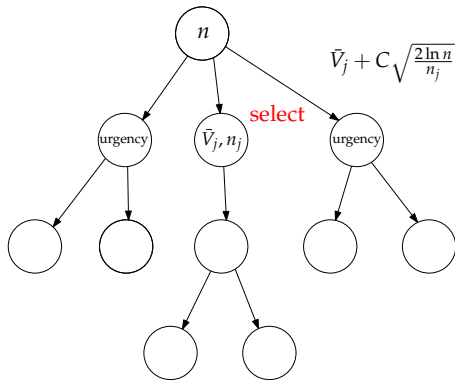
UCT Monte Carlo Tree Search XI

(Browne 2012; Browne et al. 2012; Bradberry 2015)



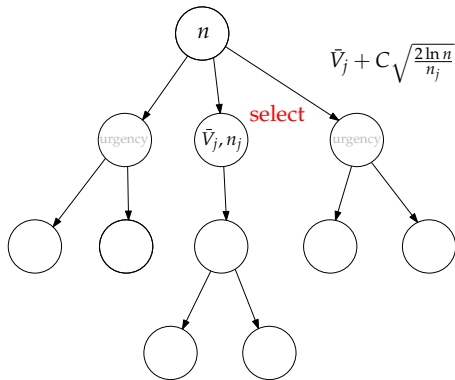
UCT Monte Carlo Tree Search XII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



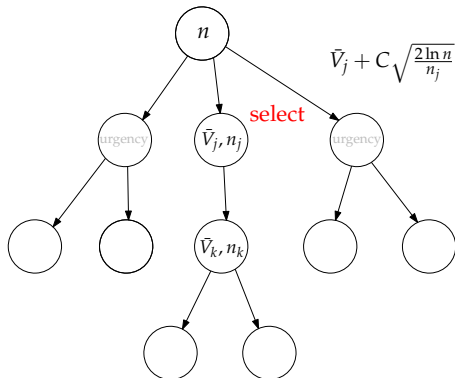
UCT Monte Carlo Tree Search XIII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



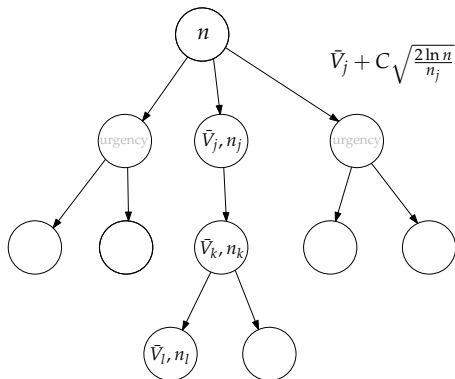
UCT Monte Carlo Tree Search XIV

(Browne 2012; Browne et al. 2012; Bradberry 2015)



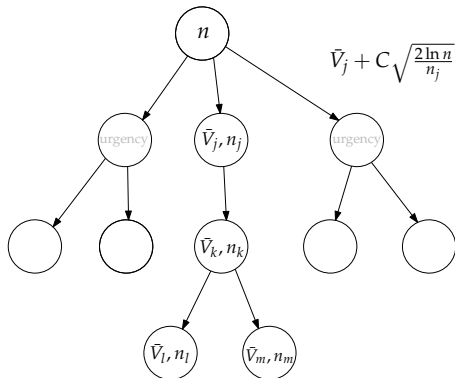
UCT Monte Carlo Tree Search XV

(Browne 2012; Browne et al. 2012; Bradberry 2015)



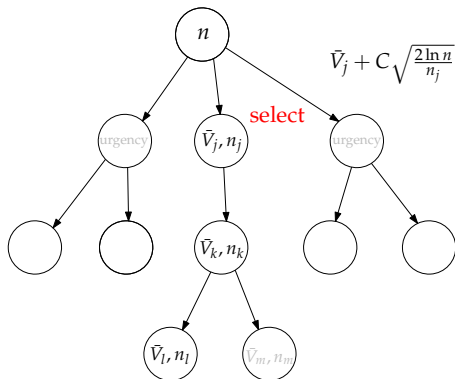
UCT Monte Carlo Tree Search XVI

(Browne 2012; Browne et al. 2012; Bradberry 2015)



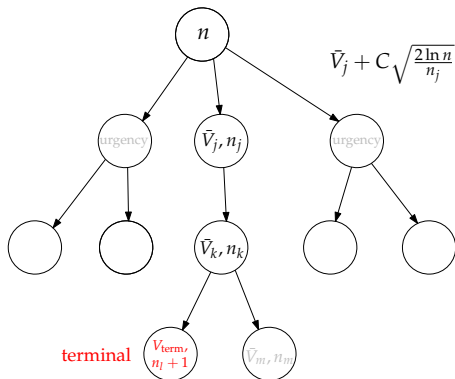
UCT Monte Carlo Tree Search XVII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



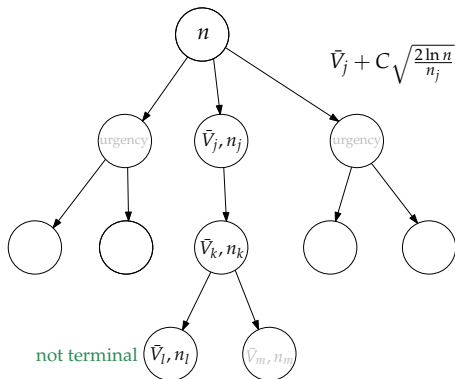
UCT Monte Carlo Tree Search XVIII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



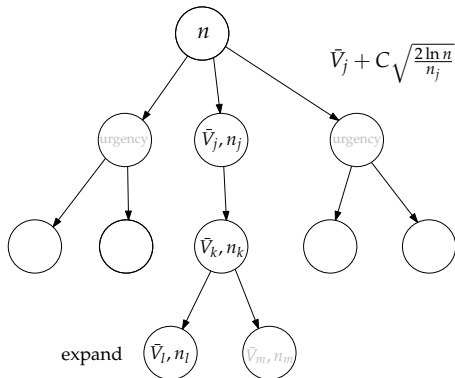
UCT Monte Carlo Tree Search XIX

(Browne 2012; Browne et al. 2012; Bradberry 2015)



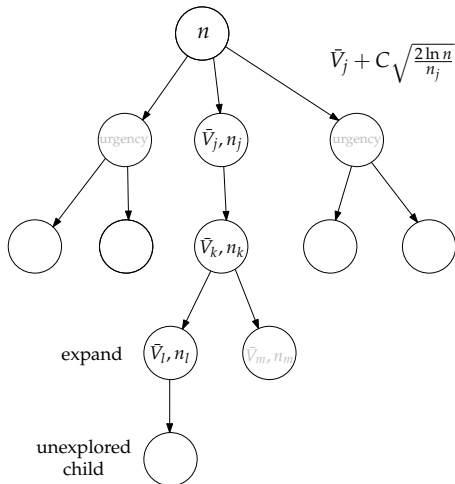
UCT Monte Carlo Tree Search XX

(Browne 2012; Browne et al. 2012; Bradberry 2015)



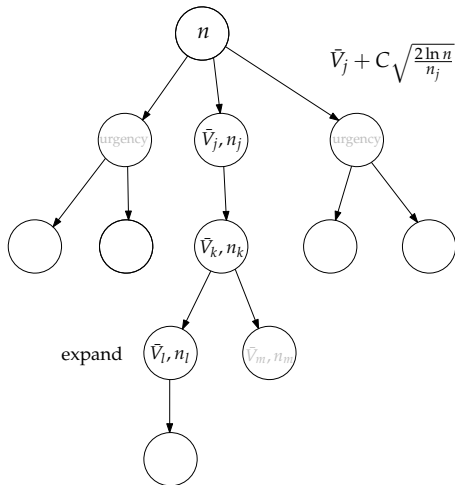
UCT Monte Carlo Tree Search XXI

(Browne 2012; Browne et al. 2012; Bradberry 2015)



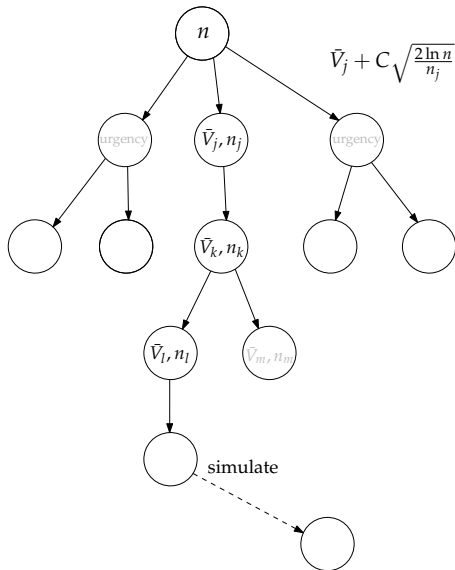
UCT Monte Carlo Tree Search XXII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



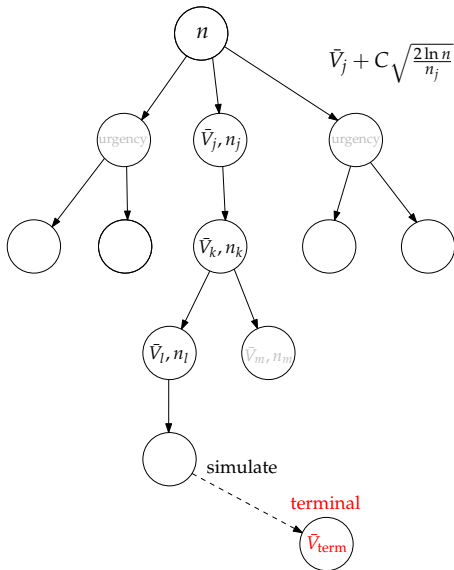
UCT Monte Carlo Tree Search XXIII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



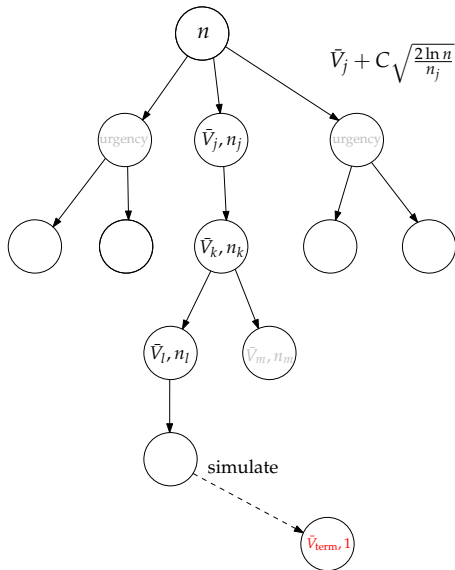
UCT Monte Carlo Tree Search XXIV

(Browne 2012; Browne et al. 2012; Bradberry 2015)



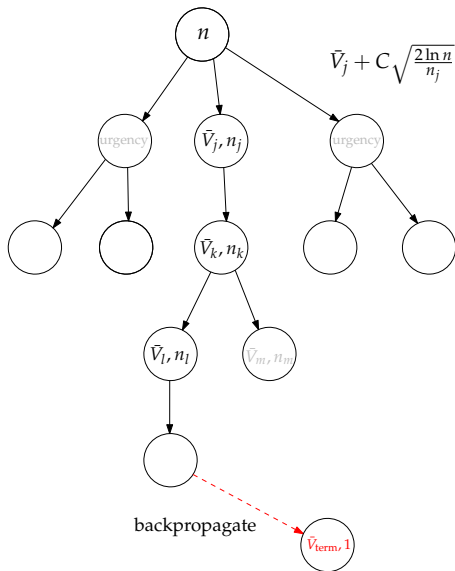
UCT Monte Carlo Tree Search XXV

(Browne 2012; Browne et al. 2012; Bradberry 2015)



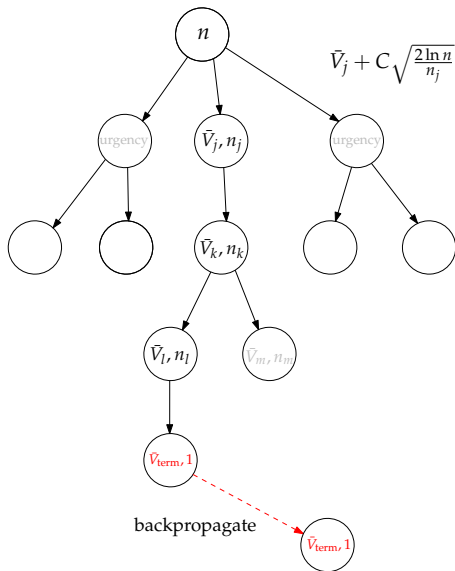
UCT Monte Carlo Tree Search XXVI

(Browne 2012; Browne et al. 2012; Bradberry 2015)



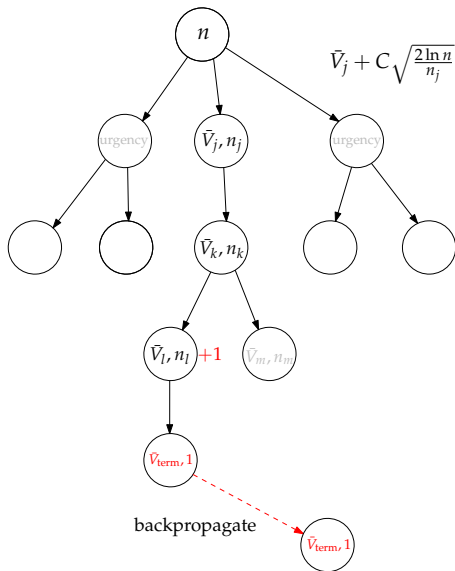
UCT Monte Carlo Tree Search XXVII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



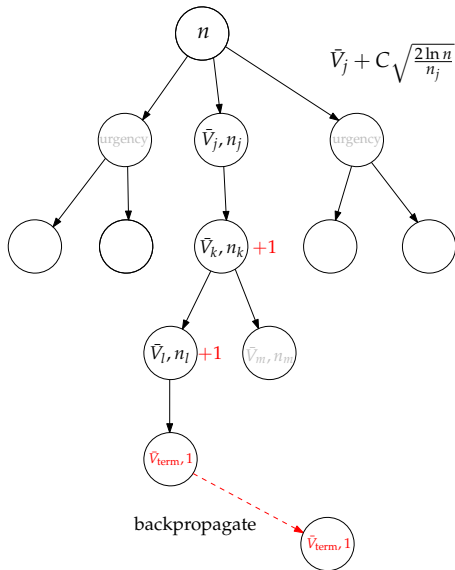
UCT Monte Carlo Tree Search XXVIII

(Browne 2012; Browne et al. 2012; Bradberry 2015)



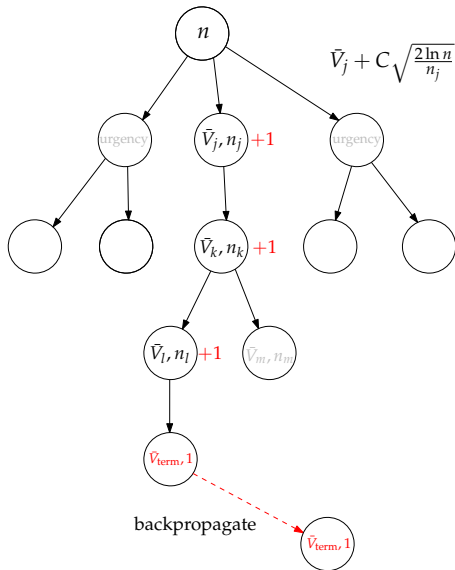
UCT Monte Carlo Tree Search XXIX

(Browne 2012; Browne et al. 2012; Bradberry 2015)



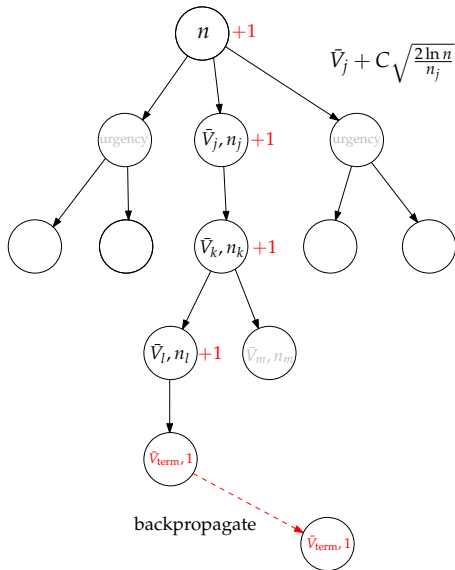
UCT Monte Carlo Tree Search XXX

(Browne 2012; Browne et al. 2012; Bradberry 2015)

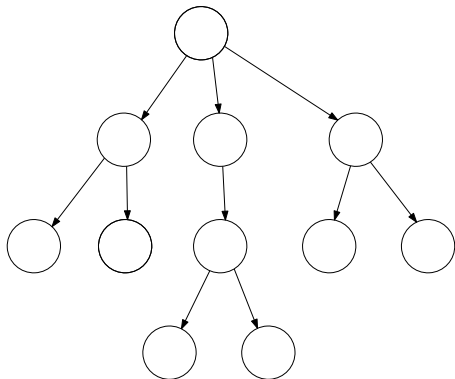


UCT Monte Carlo Tree Search XXXI

(Browne 2012; Browne et al. 2012; Bradberry 2015)

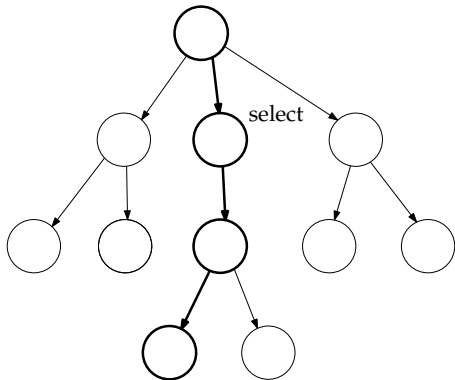


Summary



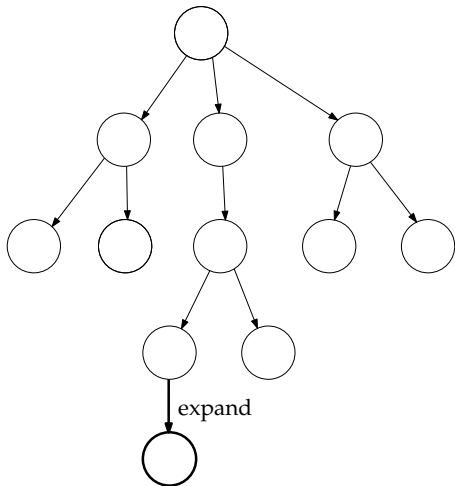
Summary

1 select



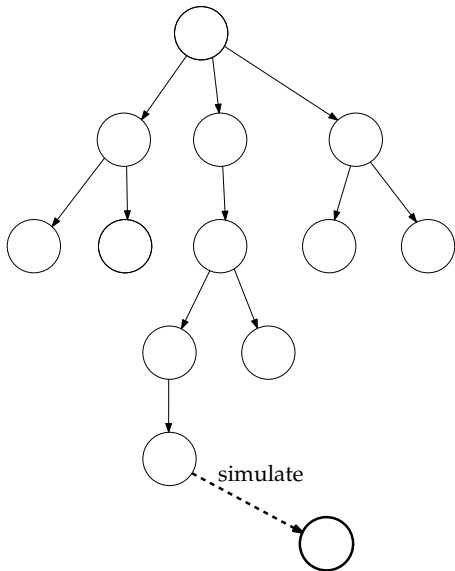
Summary

- 1 select
- 2 expand



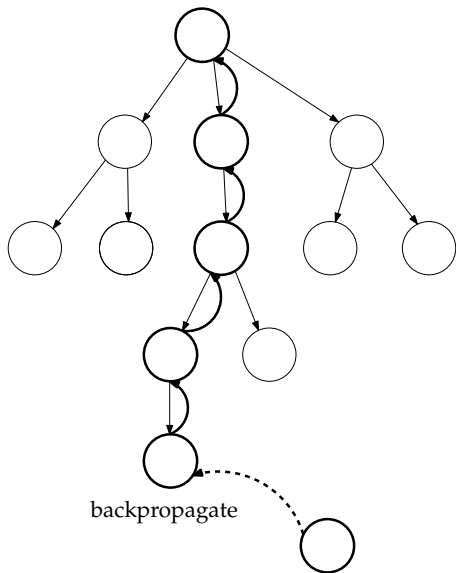
Summary

- 1 select
- 2 expand
- 3 simulate



Summary

- 1 select
- 2 expand
- 3 simulate
- 4 backpropagate



Note:

- we treated it as a puzzle problem
- rewards just positive

But:

- in a game, antagonistic situation
- either: use NEG-MAX picture
- turn reward around at each step (multiply by -1 for each level)

(Browne 2012)

- or: have utility for the player of the particular incremented if they won the game

Mystery Factor: Urgency

Confidence Bound

- consider a sequence of random rewards (value payoffs)
- with mean \bar{V}
- it is not perfectly accurate
- from Hoeffding's inequality (google it if you dare!), one gets that the true mean is “with good probability” in an interval

$$\left[\bar{V}_j - \sqrt{\frac{2 \ln n}{n_j}}, \bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}} \right]$$

if option j is visited n_j times and n total runs have been made

- it can be shown that selecting the branch with highest **upper confidence bound** (UCB)

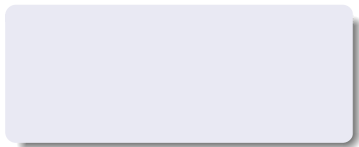
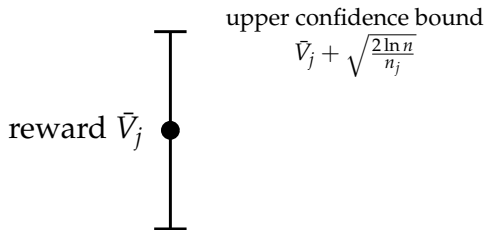
$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

minimizes regret asymptotically

(Auer 2003; Kocsis and Szepesvári 2006)

Criterion

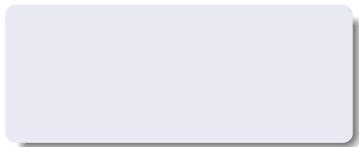
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

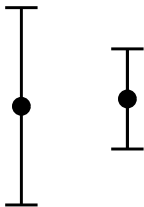
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

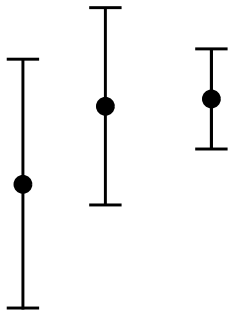
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

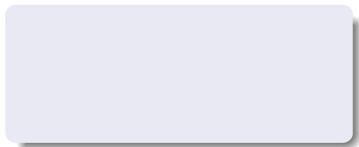
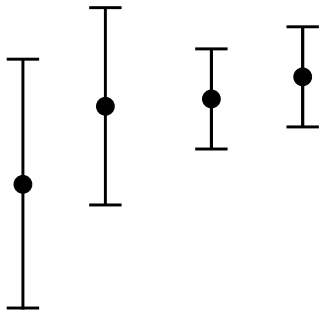
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

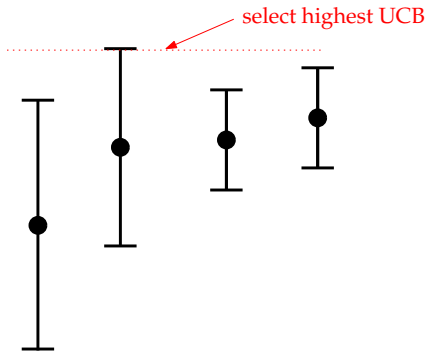
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

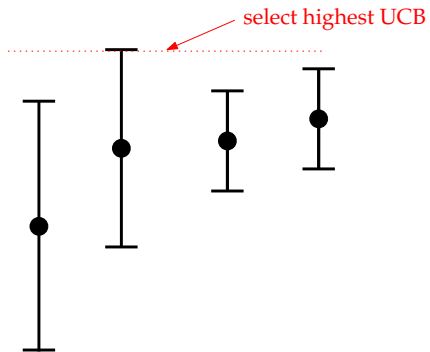
(Browne 2012; Browne et al. 2012)



$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

(Browne 2012; Browne et al. 2012)

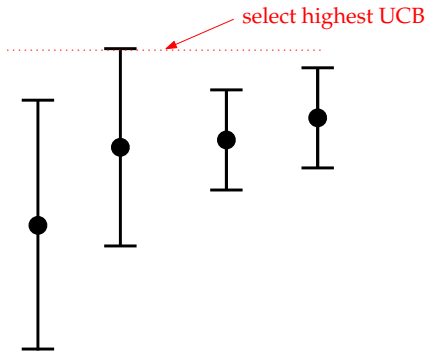


- highest UCB

$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

(Browne 2012; Browne et al. 2012)

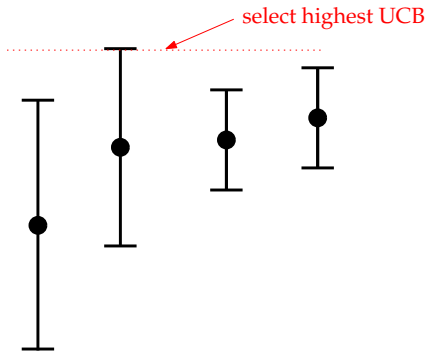


- highest UCB
- **not** highest reward

$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Criterion

(Browne 2012; Browne et al. 2012)



- highest UCB
- **not** highest reward
- **not** widest spread

$$\bar{V}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

UCT Pseudocode

(Browne 2012)

The above was taken directly from Cameron Browne slides.