1 handout: slides

UCT

Monte Carlo Tree Search (MCTS)

on-line action selection in MDPs

descent using current scores
roll-out Monte Carlo until termination
update scores in tree
growth add node to tree

easy to parallelize

Upper Confidence Bounds on Trees (UCT, ECML 2006)

greedy is a poor strategy
be humble — recognize your uncertainty!
popular strategy: "optimism in the face of uncertainty"

$$W(s,a) = \text{total reward}$$
$$N(s,a) = \text{number of times tried}$$
$$Z(s,a) = \frac{W(s,a)}{N(s,a)} + C\sqrt{\frac{\log N(s)}{N(s,a)}}$$

Using UCT

roll-out policy must be fast and good but not deterministic
lots of extensions proposed
in MoGo, $C = 0$ (although MC-RAVE provided pseudo-exploration)

Break

- bring device with email/web access on Thursday for evaluations
- WI papers by Friday at 2pm

AlphaGo

Go

- branching factor: probabilistic beam
- search depth: UCT
- static evaluation: deep neural net

fast policy net $p_\pi(a|s)$ fast approximation for roll-outs

24% accurate, linear, 2μs

SL policy net $p_\sigma(a|s)$ to reduce branching factor

trained from 30M positions and expert moves

57% accuracy, 13 layers, 3ms

RL policy net $p_\rho(a|s)$ improves SL by self-play for 3 weeks

beats SL 80%, beats Pachi 85%

value net $v_\theta(s)$ predicts RL outcome from state

trained from 30M positions from separate games for week

almost as good as rollouts with RL but 15,000 times faster

Tree Search in AlphaGo

$$\operatorname*{argmax}_a Q(S,a) + c \cdot P(s,a) \frac{\sqrt{\sum_b N_r(s,b)}}{1 + N_r(s,a)}$$

$P(s,a)$ given by SL policy net (not RL!)
N_r = number of rollouts

at leaves:

$$v(s_L) = (1-\lambda) v_\theta(s_L) + \lambda z_L$$

where z_L is a rollout using fast policy p_π
in tree:

$$Q(s,a) = (1-\lambda)\frac{W_v(s,a)}{N_v(s,a)} + \lambda \frac{W_r(s,a)}{N_r(s,a)}$$

tree reused after move. search continues during opponent's move.

EOLQs

- What question didn't you get to ask today?
- What's still confusing?
- What would you like to hear more about?

Please write down your most pressing question about AI and put it in the box on your way out.
Thanks!