

アピール文書



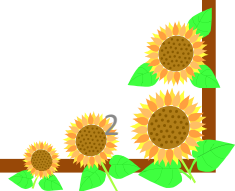
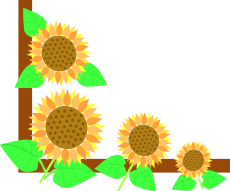
ひまわり 開発者
山本 一将



探索部



- Null Move Pruning
- Killer Move
- SEE
- PVS
- History Heuristic
- LMR
- Futurity Pruning
- Aspirationサーチ
- 並列探索
- 1手詰め
- 3手詰め

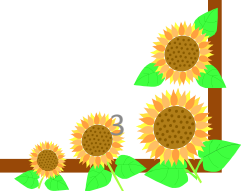
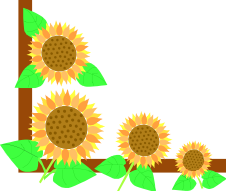




評価部



- 特徴ベクトルは
Bonanza方式の駒割＋KPP＋KKPを使用
- **学習を方策勾配法によって行っている点が特徴**
(Bonanzaメソッドは使用していない)





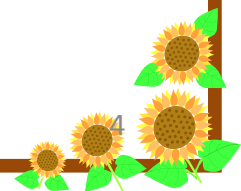
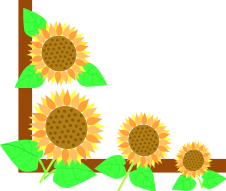
方策勾配法について



- 強化学習の一手法
- 報酬の設定が柔軟に行える。例えば・・・
 - 勝ったらプラスの報酬、負けたらマイナスの報酬
 - プロの棋譜と一致したら報酬を与える
 - 入玉したら報酬を与える

など

- 報酬の設定次第ではプロの棋譜を使用せずに良い評価関数を作成できると考えています。

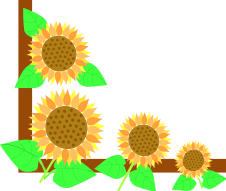




ひまわりの学習



- プロの指し手と一致
 - プロの棋譜棋譜5万8千局を使用
(方策勾配法の教師付き学習については後述)
- 深い探索を教師とする学習
 - プロの棋譜を用いない学習
 - Bootstrap的な学習を方策勾配法で行う





方策勾配法の教師付き学習



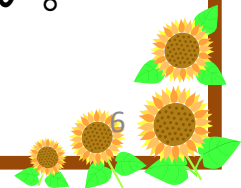
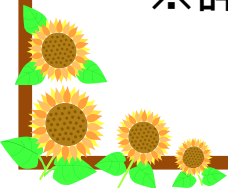
局面 s におけるプロ棋士の着手を a とする。
この時の学習式は以下のとおり。

$$\Delta\omega = \varepsilon \cdot \frac{1}{T} \left(\frac{\partial E(a, s; \omega)}{\partial \omega} - \sum_{x \in A} \pi(x | s) \frac{\partial E(x, s; \omega)}{\partial \omega} \right)$$

ただし、

$$\pi(a | s) = \frac{\exp(E(a, s; \omega)/T)}{Z} \quad Z \equiv \sum_{x \in A} \exp\left(\frac{E(x, s; \omega)}{T}\right)$$

※詳しく知りたい方は、前述の参考文献を参照するか、直接聞くなどして下さい。





来年までに頑張りたいこと



- 方策勾配法による学習の工夫
 - 序盤・中盤・終盤をそれぞれ学習するなど
 - プロの棋譜はなるべく使わない方向にしてい
 - 目標は新手を編み出すこと！
- 探索部のStockfish化

