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
- Futirity Pruning

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。

ひまわりの報酬設定

- プロの指し手と一致
  - 比較的最近の棋譜2万局を使用
    （方策勾配法の教師付き学習については後述）

- 自己対局の勝敗
  - 勝ったら＋１、負けたら－１を報酬として設定
  - 参考文献：http://ci.nii.ac.jp/naid/110009579798

方策勾配法の教師付き学習（1／2）

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。

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

ただし、

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

※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。

方策勾配法の教師付き学習（2／2）

- 激指の学習手法を一般化した式になる
  - この事実が方策勾配法の報酬設定を正しく行えばプロの棋譜を使わずに強くすることが出来ると信じる理由です。

- 激指の学習より計算量が多いため、時間がかかる
  - その分精度が高いと信じたいのですが・・・

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く
  - 目標は新手を編み出すこと！
- 探索部の強化