GA将？ アピール文書

2019年3月16日 森岡 祐一

概要

- GA将は私（森岡）が趣味で作成しているコンピュータ将棋プログラムです。
- “GA将”の読み方は“がしょう”です。
- 初期バージョンの学習ルーチンにGA（遺伝的アルゴリズム）を使用していたので、この名前にしました。が、現在はGAは全く使っていません。
- 昨年はVer.9だったので“GA将!!!!!!!!”でしたが、今年はVer.10なので“GA将？”になりました。‘？’１個で‘！’１０個分という表記方法です。
- なお、本アピール文書は読み易さを重視して記載していますので、一部の表現には厳密さが欠けている事が有ります。ご了承下さい。

特徴

- 自己対局の結果から、強化学習[1]を用いてゼロから自力で学習可能です。
- 探索ルーチンの並列化はせず、シングルスレッド探索のクライアント8つで多数決合議[2]を行います。
- 合議の票の割れ方に応じて、思考時間の延長をしています。
- 定跡は一切無し。

探索ルーチン

- αβ探索、全幅ベースで各種枝刈りを実装。
- 静止探索はKFEnd流の2段階のもの[3]。ただし、生成する手は「駒を取る手と駒が成る手」に変更しています。
- 正直、ここは特筆すべき事は有りません。

評価関数

- 以下の評価項目からなる、線形の評価関数です。
  - 駒割
  - PPT（二駒関係＋手番）
  - 王将の移動可能範囲[4]
- 局面評価とパラメータ修正で同じ制御構文を書かずにすむ様に、C++のtemplateを使用して、実装を工夫してあります[5]。

合議

- 多数決合議を採用しています。
- 評価関数パラメータは、後述する学習ルーチンを用いて生成した1つのパラメータベクトルをベースにします。
- このパラメータベクトルを、Deep Learningで用いられるDropout[6]の様に「一部のパラメータをランダムに0にする」事で、複数のバリエーションのパラメータベクトルを生成します。
- 合議クライアント数は、通常探索8クライアント＋現局面での詰将棋ルーチン1クライアントです。

思考時間制御

- 前述した様に、GA将？では多数決合議を使用しています。
- そこで、「合議クライアントの票の割れ具合」をベースに思考時間の延長制御を行っています。
- 例えば、「手Aに投票したクライアントが7個で、手Bに投票したクライアントが1個」の場合は、「おそらく手AでOKだろう」と判断して、そこで思考を打ち切ります。
- 逆に、クライアントごとに投票した手が割れた場合は、現局面は難しい局面だと判断して、思考時間を延長します。

学習ルーチン（１：基本）

- GA将？の一番の売りです。
- “PGLeaf Drei”と命名した、強化学習とαβ探索を組み合わせた手法です。
- 基本となるアルゴリズムはPGQ[7]で、αβ探索のPV Leafの評価値を元に勾配計算を行う様に拡張しました。
- PGQとは方策勾配法とQ学習を組み合わせたアルゴリズムで、各アルゴリズム単体よりも良い性能を叩き出す事が、実験的に示されています。

学習ルーチン（2：拡張）

- Noisy Networks[8]風に各パラメータに「平均と標準偏差」を持たせ、その両方を学習します。
- これにより、自己対局時の棋譜のバリエーションが増え、学習効率が向上しました。

学習ルーチン（3：学習の流れ）

1. 評価関数パラメータを、極小さな乱数で初期化する。
2. 100局自己対局を行い、その結果からPGLeaf Dreiで評価関数パラメータを修正する処理を、10回繰り返す。
3. エース評価関数（過去最強のパラメータの評価関数）相手に先後入れ替えて計400局対局を行い、勝率が55%以上であれば、現時点でのパラメータをエース評価関数のパラメータとする。
4. 2に戻る。

学習ルーチン（4：アルゴリズムの概要-1）

- まず、自己対局時はSoftmax方策を用いて、ランダムに指し手を決定します。ただし、完全にランダムでは無く「良さそうな手ほど高確率で選択する」様になっています。
- 方策勾配法では、勝った時は「本譜の手は良い手だった」、負けた時は「本譜の手は悪い手だった」と判断し、評価関数パラメータを修正します。
- 例えば、勝った場合は本譜の手の評価値が上がり、それ以外の手の評価値が下がる方向に、評価関数パラメータを修正します。
- 負けた場合は、評価関数パラメータの修正方向が逆になります。

学習ルーチン（５：アルゴリズムの概要-2）

- Q学習では、本譜の局面の評価値を、その2手先（次の自分の手番）の評価値に近付ける様にパラメータを修正します。
- 方策勾配法は「自己対局の結果から、強い（勝率の高くなる）評価関数を直接作る」事を目的としたものであり、Q学習は「自己対局の結果から、精度の高い（勝率をより良く近似する）評価関数を作り、間接的に棋力向上を実現する」事を目的とするものです。
- これら2つの学習ルーチンを組み合わせる事により、従来のPGLeaf単体よりも棋力が向上しました。

学習ルーチン（６：対局数等）

- 探索条件は全幅1手＋静止探索です。
- 24時間あたり10万局程度の対局速度です。
- おそらく、2週間～1ヶ月程度は学習を行わないと、強くならないと思われます。

参考文献（1）

[1] Richard S.Suttonほか. 強化学習. 森北出版, 2000.

[2] 伊藤 毅志 ほか. 将棋における合議アルゴリズム: 多数決による手の選択. 2011, 情報処理学会論文誌 52-11, 3030 - 3037p.

[3] 有岡 雅章. 松原 仁編. コンピュータ将棋の進歩4: 将棋プログラムKFEndにおける探索. 共立出版, 2003, 18-40p.

[4] https://gasyou.hatenablog.jp/entry/20100922/1285152080

参考文献（2）

[5] https://gasyou.hatenablog.jp/entry/20120508/1336489213

[6] Nitish Srivastavaほか. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. 2014, Journal of Machine Learning Research 15, 1929-1958p.

[7] https://arxiv.org/abs/1611.01626

[8] https://arxiv.org/abs/1706.10295

最後に

- SPECIAL THANKS
  - 強化学習関連のご指導：mooopanさん（@mooopan）
  - ディスカッション等：桜丸@mEssiah_β1さん（@sakuramaru7777）
- 宣伝など。興味を持って頂けたら、御覧下さい。
  - Blog：https://gasyou.hatenablog.jp/
  - Twitter：@Gasyou_Official、@MoriokaYuichi
  - Web Site：http://gasyou.is-mine.net/