GA 将!!!!!!!!! 第 3 回将棋電王トーナメント PR 文書

2015 年 9 月 27 日 森岡 祐一

1. はじめに

- GA 将!!!!!!!!!は私（森岡）がフルスクラッチで作成したコンピュータ将棋ソフトです。
- 開発の方針としては、評価関数の精度向上を最優先目標としています。
- 評価関数のパラメータ学習には、強化学習という手法を用います（詳しくは後述）。
- 最初のバージョンでは GA（遺伝的アルゴリズム）を用いて学習を行っていた為、この様な名前にしました。
- 「GA 将」の読み方は「がしょう」です。名前の最後の「!」はバージョン番号です。
- 2015 年 5 月の第 25 回世界コンピュータ将棋選手権から、棋力向上はほとんどありません。多少なりとも改良してトーナメントに臨む予定ですが、多分弱いままです。

2. 目標

- 棋譜や定跡といったエキスパート（≒プロ棋士）の知識データを用いずに、ほとんど事前知識が無い状態から学習を行い、その結果どのような棋風になるかを知りたいと思って、GA 将!!!!!!!!!の開発を始めました。
- 例えば「軽く囲ってから急戦を仕掛けるのがベスト」という結果になり新しい戦法を作り出せるのか、あるいは「矢倉や穴熊に囲うのが良い」という人間の経験に追従する結果になるのか、それが知りたいです。
- ただし、GA 将!!!!!!!!!のプログラム（主に評価関数と探索ルーチン）には、私が棋書を読むなどして得た、将棋固有の知識が含まれています。ですので、将棋の知識が完全に 0 の状態から学習を行う訳ではありません。
- 余談ですが、多層ニューラルネットワークや SVM（SVR）、ディープラーニングといったアルゴリズムを用いれば、将棋の知識を全く使用せずにコンピュータ将棋を作成出来る可能性があります。ただ、私は評価関数・探索ルーチンに関してはコンピュータ将棋専用のものを用いるのがベストだと考えています。

3. 探索ルーチン

- 探索ルーチンは凝った事をせず、シンプルな実装に留めています。
  - $\alpha\beta$ 探索で全幅ベースの探索部＋KFEnd 流の二段階静止探索。
  - ムーブオーダリングは置換表＋History Heuristic のみ。
  - 全幅ベース部では LMR・Null-Move Pruning・Extended Futility Pruning・Move Count Based Pruning での枝刈りあり。
  - 王手は 0.5 手延長。
- マルチスレッド探索と Ponder（予測読み）も一応実装しました。
- 探索のスピードは Core i7 5960X を搭載した PC で 3Mnps 前後（マルチスレッド探索時。局面によって変動します）。1 秒で序盤 13 手、終盤 9 手程度しか読めません。

4. 評価関数

- 以下の特徴量の線形和で評価値を決める、コンピュータ将棋用としては比較的普通の作りになっています。
  - 駒割の評価
  - 駒の絶対位置の評価
  - 二つの駒の絶対位置関係の評価
  - 三つの駒の相対位置関係の評価
  - 王将の移動可能範囲の評価
  - 王将周辺の利きの数の評価

- パラメータ数は上記全てで 860 万個程です。

5. 機械学習・強化学習・SR-PGLeaf について

5.1 なぜ「学習」が必要か？

- 近年のコンピュータ将棋は、数千万～数億個の評価関数パラメータを持つソフトが多く存在します。
- しかし、それらのパラメータ全てに、手作業で適切な値を設定するのは現実的ではありません。
- そこで、「機械学習」によりパラメータに値を設定するのが近年の主流です。
- （実際には、コンピュータ将棋における機械学習の実用化により、多数のパラメータを持つ評価関数が使用可能になったと言う方が正確ですが。）

5.2 教師あり学習と強化学習の違い

- 「機械学習」とは「従来人手で行っていた作業を、機械（コンピュータ）が実行出来る様にする為の技術・手法」です。
- 機械学習は大きく分けると「教師あり学習」「強化学習」「教師なし学習」の 3 つの分野からなります。ただし、教師なし学習はコンピュータ将棋への応用例は聞いた事が無いので、ここでの説明は割愛します。
- 教師あり学習は、文字通り「教師」が存在する学習方法です。例えばコンピュータ将棋の場合であれば「この局面ではこの手が最善手である」とか「この局面の評価値は何点である」といった情報を与えるのが「教師」です。
- 強化学習では「教師」は存在せず、かわりに「報酬」と呼ばれる信号を頼りに学習を行います。コンピュータの行った一連の行動（将棋の場合であれば指し手選択）が「良い」ものであればプラスの、「悪い」ものであればマイナスの報酬が与えられます。
- ただし、報酬には遅延やノイズがあります。例えば、良い手を指した直後にプラスの報酬が与えられるとは限りませんし（遅延）、悪手を指した後にプラスの報酬が与えられる事もあります（ノイズ）。その為、何らかの方法で自分（コンピュータ）の指した手の良し悪しを判断する必要があります。

5.3 Bonanza Method と SR-PGLeaf

- Bonanza Method ではエキスパート（プロ棋士や上位のアマチュア等）の棋譜を大量に用意し、棋譜に含まれる局面での棋譜の指し手を「教師」として学習します。
- これは、人間が将棋を学習する際の「棋譜並べ」に類似した学習手法です。
- それに対して、SR-PGLeaf では自己対局の終局時に「勝ちなら+1、負けなら-1、引き分けなら 0」の報酬を与え、報酬を最大化する様に（＝勝率を上げられる様に）学習を行います。
- つまり、人間が実戦を通して学習するのと似た方法で学習を行います。

5.4 SR-PGLeaf の詳細

- SR-PGLeaf(Split Reward-Policy Gradient with Leaf)は、強化学習の既存手法である方策勾配法(Policy Gradient Method)をベースとして、指し手選択・パラメータ修正部分にα β 探索を組み合わせたものです。
- SR-PGLeaf のベースとなっている PGLeaf の詳細は、GPW での発表資料に記載してあります。詳しく知りたい方は下記 PDF を参照して下さい。

http://gasyou.is-mine.net/archive/GPW2012_P-3.pdf

- 前述しましたが、強化学習では「正解(局面での指し手や評価値)」は与えられません。その代わり、例えば終局時に勝敗に応じた「報酬」を与えられ、その報酬をより多く得る為にはどうすれば良いかを試行錯誤しながら学習を行います。
- PGLeaf では終局時に勝敗に応じた報酬を与え、これを手掛かりに学習を行います。この為、例えば「中盤まで優勢に進めていたが、終盤で逆転負けした」という場合に、「序盤～中盤では良い手を指していた」とプログラムが判断出来ません。
- そこで、SR-PGLeaf では自己対局中の評価値の変化を元に、指し手の良し悪しを判断させる事にしました。
- ここで、以下の通り用語を定義します。
  - 確定報酬:終局時に与えられる報酬。勝ちなら+1、負けなら-1、引き分けなら 0。
  - 予測報酬:ある局面から、指し手にある程度のランダムさを与えた上で十分な数の対局を行った際に得られる、確定報酬の期待値。ただし、終局時の予測報酬は確定報酬と同じ値とする。
  - 分割報酬:ある局面 P より先の局面の予測報酬と、局面 P での予測報酬の差。
- 予測報酬ですが、局面の評価値とシグモイド関数を用いて「2 * sigmoid(評価値) - 1」で近似します。シグモイド関数のゲインは、対局中の評価値とその対局の確定報酬を全て保存しておき、一定間隔で最適な値に設定します。
- PGLeaf と SR-PGLeaf の違いは下記の 2 点です。
  1. SR-PGLeaf では、1 回の指し手選択を 1 エピソードとして扱う(1 手ごとに指し手の良し悪しを判断する)。
  2. 1 エピソード終了ごとに分割報酬を強化学習エージェントに与える。
- 上記の修正により、対局結果にかかわらず評価値の上昇した(＝分割報酬がプラスになった)手は良い手だと判断します。同様に、評価値の下降した手は悪手と判断します。

5.5 GA 将!!!!!!!!!!における強化学習の適用

- SR-PGLeaf では自己対局や他のソフトとの対局の結果から学習が可能なので、既存手法と異なり棋譜・定跡等の人間の知識データを必要とせずに学習が可能です。
- GA 将!!!!!!!!!における SR-PGLeaf では、自己対局した結果から評価値が上昇した指し手は正例（良い手）、評価値が下がった指し手は負例（悪い手）と判断してパラメータを修正します。
- GA 将!!!!!!!!!では自己対局を何十～何百万局も行い、その結果をベースに学習（評価関数のパラメータ修正）を行います。
- 自己対局・パラメータ修正時には、先手と後手で同じ評価関数を使用します。
- つまり、人間が自分一人で対局を行い、その結果から学習する様な感じです。
- 大体 50 万～100 万局程度で学習が収束します。全幅 3 手＋静止探索 6 手で自己対局すると、24 時間あたり 5 万局前後の速度なので、学習が完了するまで 2～4 週間程度かかります。（Core 5960X 8 コアマシン使用時。）

6. その他

- 定跡は一切使用していません。なので、初手から延々数十秒考えます。
- 数十秒消費した挙句に、単に飛車先の歩を突くだけとか角道を開けるだけとか、そういう風にもなりますので、序盤に関しては自己対局の棋譜から定跡を構築する等の対策をしたいところです。
- 詰将棋ルーチンは ABC 探索をベースにした簡単なものを実装しています。10 秒前後の思考時間があれば、20～30 手程度の詰みを見付ける事は可能です。なので、実戦用の詰将棋ルーチンとしては必要最小限の機能・速度はあります。

7. 棋力

- 2015 年 3 月 29 日現在、floodgate の 2 週間レーティングが 2219 です（Athlon 5350 搭載 PC での結果）。おそらく一次予選突破のボーダーライン上にいるものと思われます。
- 他の思考エンジンとの対局結果ですが、対 ssp（プチ将棋付属の思考エンジン）では勝率 92%前後、対 Bonanza 6.0（5 手読みに制限）では勝率 59%前後です。

8. おわりに

- ◆ 以前の選手権では5八玉戦法（勝手に命名）で二次予選に進出したりと、色々と見ていて面白い将棋を指していました。
- ◆ 今回も観戦者の方に楽しんで頂ける様、頑張って開発してきます。
- ◆ ブログ・Twitter 等もやってます。GA 将!!!!!!!!!に興味を持って頂ければ、こちらの方もどうぞよろしくお願いします。

http://d.hatena.ne.jp/Gasyou

https://twitter.com/MoriokaYuichi

http://gasyou.is-mine.net/