

「技巧」詳細アピール文書

2024年5月19日

出村 洋介

1 独自に工夫した点

1.1 Gumbel AlphaZero を用いた効率的な強化学習

AlphaZero [1]の学習を効率化した Gumbel AlphaZero [2]を用いて、既存の棋譜を使わずに強化学習を行いました。Gumbel AlphaZero のアルゴリズムを用いたことにより、少ないシミュレーション数で計算時間を抑えた効率的な強化学習が可能となりました。

今年の技巧では、乱数で初期化されたネットワークから約 5000 万対局の自己対局で学習を行いました。学習時間は 1 週間程度に抑えられています（学習には NVIDIA 社の GeForce RTX 2080 Ti~RTX 4090 クラスの GPU を 20 台程度利用）。

1.2 強化学習時の初期局面を多様化

自己対局時に様々な局面を積極的に経験させるため、学習時の自己対局で用いる初期局面の多様化を行ないました。具体的には、平手初期局面から 2 手ランダムに指した局面（下図(b)）や、駒の配置を一定の制約のもと並べ替えた初期局面（下図(c)、「チェス 960」にならない「将棋 81 万」と呼んでいます）を初期局面として強化学習を行いました [3]。



(a) 平手初期局面



(b) 2手ランダムの初期局面(例)



(c) 将棋81万の初期局面(例)

1.3 ネットワークの改良

推論に用いるネットワークは、Gumbel AlphaZero 論文 [2]で用いられた bottleneck blocks をベースに独自の改良を行いました。（この改良については別の機会に発表させていただきます。）

2 開発動機

深層強化学習を用いた将棋 AI の作成には大規模な計算リソースが必要という状況を少しでも改善できればと考え、今年の技巧では、各種効率化により計算資源を抑えつつ、既存の棋譜を使わずにゼロから強化学習を行うことを目指して開発を行いました。また、既存の棋譜や将棋用のライブラリを利用しなかったのは、コンピュータの自己対局のみで学習するどのような将棋を指すのか純粋に見てみたいという気持ちもありました。

3 開発過程

2016年の選手権後にC++版の技巧を公開していましたが、新たなプログラミング言語に挑戦したいと考え、心機一転 Rust と Python を使ってフルスクラッチで書き直しました。

今回の開発では、計算コストを下げた実験を多数行い開発サイクルを速くして各種改良点を地道に探しました。普段の実験では、棋譜数を1000万局程度とし、さらにネットワークのチャンネル数やブロック数を減らすなどして、教師あり学習ならば1GPUで1日程度、強化学習なら10GPU程度で1日程度で1回の実験が終わる計算量に抑えていました。

4 実験結果

選手権に向けては、学習棋譜数は5000万対局程度まで増やし、ほぼ飽和状態となるまで自己対局による強化学習を行いました。自己対局のシミュレーション数は、概ね1000万対局まで1手32シミュレーション、1000万~4000万対局まで64シミュレーション、4000万対局以降は128シミュレーションと段階的に増加させています。(今回は学習途中で手動で学習条件を調整したため実際の対局数には若干誤差が出ています。)

今年の決勝リーグで用いたネットワークと探索部の実験結果は以下のとおりです。

技巧 (100 ノード) 対 水匠 5 (40,000 ノード) ¹	+202 Elo
技巧 (800 ノード) 対 水匠 5 (250,000 ノード) ¹	+127 Elo
初期局面の探索速度 (15 秒間の平均 NPS)	100,153 局面/秒 (RTX 4090)
	58,493 局面/秒 (RTX 3090)

5 追試可能か

強化学習ではランダムな要素が複数あるため(ネットワークの重み初期化、自己対局時の開始局面の選択、自己対局時の指し手など)、完全な再現は難しいですが、以下の参考文献などを参考に概ね追試可能と考えています。ネットワーク関係の改良は後日別の機会に改めて発表させていただければ幸いです。

参考文献

- [1] D. Silver, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419): 1140–1144, 2018.
- [2] I. Danihelka, et al. Policy improvement by planning with Gumbel. In *International Conference on Learning Representations*, 2022.
- [3] 出村, 金子. 将棋 81 万: 強化学習のための多様性を持った将棋初期局面集. 第 28 回ゲームプログラミングワークショップ, pp. 111–118, 2023.

¹ 対局は各 3000 局。MCTS 探索のバッチサイズは 1。水匠 5 との対局実験では、公開されている dlshogi の互角局面集 (<https://tadaoyamaoka.hatenablog.com/entry/2022/12/31/114258>) を 24 手目まで利用。