Polonaise アピール文書

谷合廣紀

2024 年 5 月 15 日

1　開発動機

近年の将棋 AI 開発において、dlshogi をはじめとした AlphaZero 系のアプローチは非常に強力な手法として注目されています。しかしながら、将棋の盤面を画像として捉え、ResNet などの畳み込みニューラルネットワーク（CNN）を用いる手法には疑問を感じていました。

CNN の特徴の一つに位置不変性があります。これは、画像中の特徴がどの位置にあっても同じ特徴量に変換される特性です。囲碁においてはこの特性が有効ですが、将棋においては必ずしも有効とはいえません。例えば、美濃囲いは２八玉・３八銀・４九金という特定の配置だからこそ囲いとしての固さを発揮します。これが２七玉・３七銀・４八金や３八玉・４八銀・５九金とタテやヨコにずれた配置では、同じ意味を持たなくなります。さらに、将棋には飛・角・香といった飛び駒があり、これらは最大で８マス離れた位置に利きを作ることができます。これを特徴量として捉えるためには、3x3 の畳み込み層では 3 層必要になります。

このように将棋において CNN は必ずしも最適なモデルとは言えません。そのため自然言語処理で大きな成果を上げている Transformer を将棋 AI に応用することに着目しました。将棋の盤面を文字列に変換し、自然言語処理モデルで学習・推論を行うことで、従来の手法とは異なる新しいアプローチを模索しました。これが、Polonaise の開発動機です。

2　開発過程

最初の実装は、YouTube チャンネル『予備校のノリで学ぶ「大学の数学・物理」』の動画に出演した際に作成した BERT-MCTS でした[*1]。動画出演のオファーを受けた際に、インパクトのあるコンテンツを提供したいという思いと、自然言語処理モデルを将棋 AI に応用するというアイデアを持っていたことから、これを良い機会と捉えて開発に着手しました。

BERT-MCTS は、実装の一部に誤りがあったり、モデルの学習が不十分だったため、あまり強力な将棋 AI にはなりませんでした。しかし、このアプローチ自体には大きな可能性を感じました。このため、さらにこのアプローチを成長させ、第 32 回世界コンピュータ将棋選手権に出場することを決めました。探索部には、ふかうら王のアルゴリズムをベースに実装しました。

第 32 回から大きなアップデートは探索部にはありませんが、主にモデルの改良を重ねてきました。そして今回の第 34 回世界コンピュータ将棋選手権では、BERT-large という大規模なモデルに挑戦しました。これ

[*1] プロ棋士自作の将棋 AI と戦ったら色々とヤバかった: https://www.youtube.com/watch?v=2Vl6Ao4GaSQ
BERT-MCTS のコード: https://github.com/nyoki-mtl/bert-mcts-youtube

が可能になったのは、多くの開発者が無料で良質な教師データを公開してくださったおかげです。また、モデルの学習には Google の TPU v3-8 を利用しています。

3　独自の工夫

3.1　モデル入力のエンコード

モデルに盤面を入力するにあたって、まずは盤面情報を数値行列である入力特徴量に変換する必要があります。dlshogi では駒の位置や利きなどを 9x9 の 2 次元行列にエンコードしていき、最終的に 9x9x 特徴数の大きさを持つ入力特徴量を得ています。この入力特徴量は CNN を使い推論されていくため、dlshogi は画像処理的なエンコードと捉えることができます。

一方の Polonaise では、盤面を 1 一から順に見ていき、 1 一、 1 二 ･･･ 9 九の駒と先後の持ち駒（7 種 x2）を並べた 95 字の文字列にエンコードすることで入力特徴量を得ます。この入力特徴量はモデルの最初の層で埋め込み層によりベクトルに変換されて推論されていくため、自然言語処理的なエンコードと捉えることができます。

3.2　モデル出力

dlshogi で採用されている policy の出力は、「着手するマス」と「その駒はどの方向から来たか」の組み合わせで表現されます。「着手するマス」は 81 マスあり、「どの方向から」は 27 通りあるため、その組み合わせは 2187 通りです。したがって policy は 2187 通りのクラス分類問題として表現されています。

しかし、「1 一のマス」に「下がる」や「左に寄る」といった動きは将棋の合法手として存在しません。このように dlshogi の policy 表現の中には決して現れない組み合わせがいくつかあります。それら非合法手を数え上げていくと 691 通りあり、約 32% が非合法手となっていることがわかります。

実験の結果、policy の出力を 2187 クラスの分類問題として解くよりも、非合法手 691 通りを除いた 1496 のクラス分類問題として解いた方が、policy の学習がうまくいくことがわかりました。そのため Polonaise では 1496 のクラス分類問題として policy の学習を行っています。

3.3　PVM ネットワーク

モデルの出力として policy, value に加えて mate_prob すなわち入力局面が詰むかどうかを出力しています。この mate_prob がある閾値 (大会では 0.5) を越えたらその局面が詰み探索キューに追加されて、待機している複数の df-pn ソルバーで詰み探索を行い、詰みと判定されたらそのノードの情報を勝ちに更新します。これによって dl 系が苦手とする終盤において、見落としが少なくなったように見えます。(実装が大会直前だったため、計測データはありません。)

4　追試可能か

公開データによる教師あり学習しか行っていないため、同じ入力エンコードとモデル構造を用いれば追試可能です。