Polonaise アピール文書

谷合廣紀

2024 年 5 月 4 日

1　独自の工夫

基本は dlshogi/ふかうら王などのいわゆる DL 系で採用されている、policy+value Network + MCST のアプローチを取っています。dlshogi/ふかうら王と大きく異なるのは、モデル構造とその入出力です。

1.1　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。

具体的に Polonaise のエンコード方法を図1の盤面を使って示します。

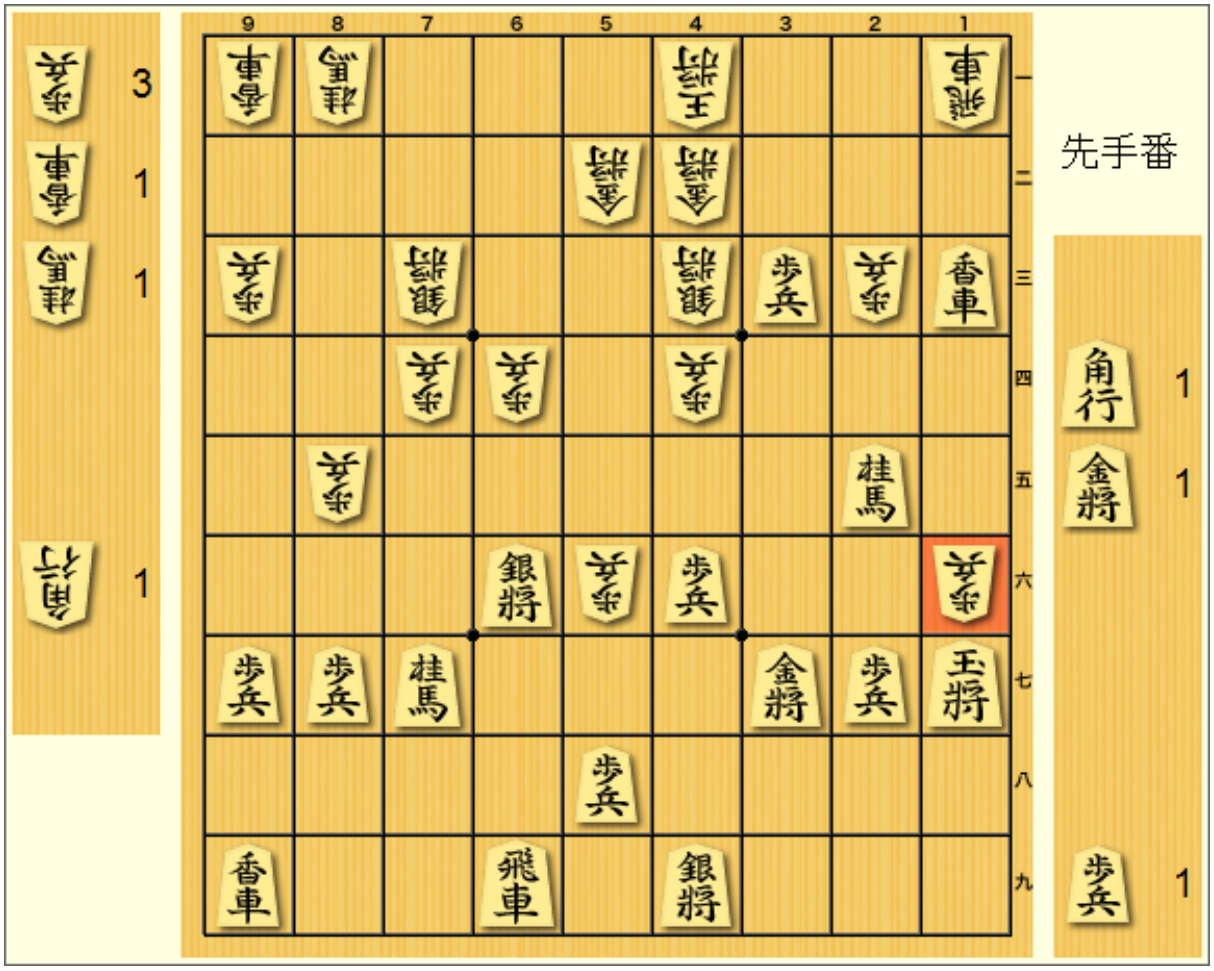

図1: 入力盤面

この盤面の駒を 1 一から順に見ていくと、「後手の飛車」、「空きマス」、「先手の香車」･･･ と続きます。また、持ち駒は先手は歩が 2 枚、金が 1 枚、角が 1 枚です。盤面 81 マスの情報はそのまま駒情報が入り、持ち駒はそれぞれの駒を何枚持っているかが入ります。したがって図1をエンコードすると、図2の文字列が得られます。

実際には文字列だと扱いづらいため、各文字が対応する数値 ID に変換されて、95 個の数値列がエンコードされた入力特徴量となります。

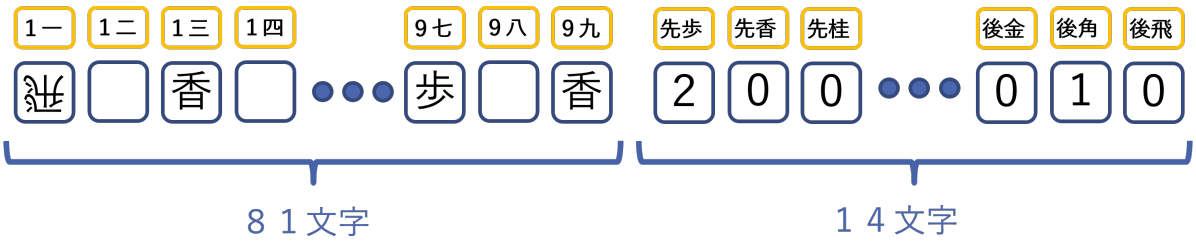

図2: 文字列エンコード

1.2　モデル構造

95 個の数値列は、最初の埋め込み層によって 95x256 の大きさの行列に変換されます。これまでは gMLP[?] と呼ばれる MLP ベースのモデルを採用していましたが、今回は BERT-large 程度の中規模なモデルを採用しています。

1.3　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。

1.4　PVM ネットワーク (5/5 追記)

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。(実装できたばかりのため、計測データはなし。)

2　使用ライブラリ・使用データ

- ふかうら王
- AobaZero 公開データ
- Suisho-10Mn 公開データ

参考文献