大将軍 詳細アピール文書

横内健一・横内靖尚

概要

昨今のコンピュータ将棋において評価関数は劇的な変化を遂げてきている。Bonanza で始まった従来の 3 駒関係をベースとした評価関数から NNUE を経て、Deep Learning を用いた評価関数も登場し、その精度の高さは GPU の性能向上と相まって飛躍的に進歩している。

大将軍は、そのようなトレンドのなか、評価関数については今更ながら従来の 3 駒関係（15 年以上前の技術）をベースに大会に参加している。

開発動機

大将軍は以前に N4 や N4S という名前で大会に参加していたことがあるが、4 駒関係を用いた評価関数を使用していた。その際の経験では、序盤はそれなりに戦えるものの、中盤以降、深い読みができず、敗戦するケースが多かった。評価関数の表現力を高め大局観の精度を高めることは大事ではあるものの、将棋の場合、最終的には詰む・詰まない、を正確に読み切ることが重要であり、具体的な手順を深く読む必要がある。そのため将棋では 4 駒関係のような重い評価関数は不向きとの考えがあった。

そこで、あえて比較的軽い 3 駒関係の評価関数と最新の探索アルゴリズムを組み合わせた場合、どの程度大会で戦うことができるかをモチベーションに大会に参加した。

開発過程

元祖の Bonanza 型の 3 駒関係を出発点として、4 駒関係などの開発を進めてきたが、これらの評価関数では、手番が評価に含まれないため、先後同一局面では評価値が 0 になってしまうという課題があった。そのため、3 駒関係をベースとして、さらに手番を考慮した 3 駒のモデル（kkpt-kpp）で大会に参加していた。昨年からは一般的な手番を考慮した 3 駒のモデル（kppt 型）とした。

使用ライブラリとその選定理由

やねうら王の選定理由

ソースコードがわかりやすく、ベースのエンジンとして使用

水匠 2,3,4,4 改の選定理由

学習の棋譜生成に使用（強いソフトで生成した棋譜を利用）

独自に工夫した点

これらは過去からの工夫で、今では常識の技術ではあるが、3 駒関係のモデルでは、評価関数の差分評価（動いた駒のみの評価値を更新）をいち早く取り入れて採用している（4 駒関係の計算では、すべての駒の配置に対して評価値を計算すると組み合わせが多く膨大な計算量が必要となるため、差分計算を取り入れると 3 駒分の計算量で済む経験を利用）。また、学習においてプロの棋譜からの学習時においてミニバッチ方式を採用した。これにより、パラメータの更新回数を相対的に増やすことができ、学習を効率よく行うことができた。

今回の新しい取り組みとして、学習対象とする局面に対し、状況に応じた学習量のパラメータ調整を試みた。やねうら王を用いた学習では、深い探索の評価値や勝敗をもとに損失を計算する。また、学習の対象とする局面は、現局面ではなく、現局面から浅い探索（静止探索）を行い、その末端の局面に対して学習を行うのがよいとされている。

しかし、深い探索の進行で学習を行うとき、深い探索の進行と浅い探索の進行が連続して一致する場合、例えば図１において、深い探索の進行が、局面Ａから局面Ｂ、局面Ｃと進行し、浅い探索の進行も同様だった場合、局面Ｃについて、３回学習が行われることとなる。これが本当によいのか疑問があった。

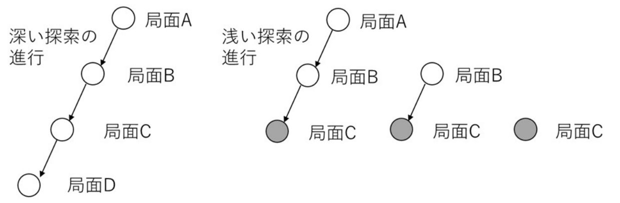

図１　深い探索と浅い探索が一致するケース

次に、図２のように、深い探索の進行と、浅い探索の進行が異なる場合に、浅い探索の末端に対して、深い探索の評価値や勝敗を反映させることがよいのかという疑問もあった。

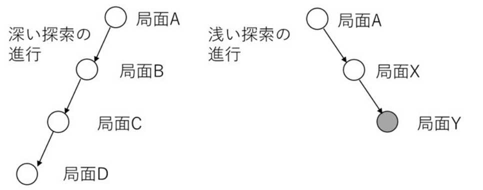

図２　深い探索と浅い探索が一致しないケース

そこで、上記のケースにおいて、学習量に係数を設けパラメータとし、調整を試みた。

実験結果

深い探索と浅い探索の進行が一致する場合の影響について、浅い探索の末端局面の学習量に対して式１の係数 factor を乗算することで調整した。この式の pv.size()は浅い探索の深さを意味する。探索が深いほど、その後に同じ局面が複数回学習される可能性が高いと仮定し、学習量を減らすことを意図している。

$$factor = \frac{p}{pv.size() + p} \qquad \text{（式１）}$$

ｐをパラメータとし、この係数を考慮せず学習したものと 1000 局対戦した勝率を図３に示す。

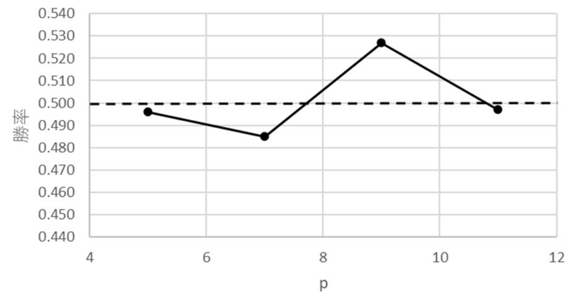

図3 パラメータpと勝率

微妙ではあるが有利な条件があり、探索の深さに応じて学習量を減らすことを採用した。

また、深い探索と浅い探索の進行が一致しないケースについては、単純に学習量に係数ｑを乗算した。係数ｑをパラメータとし、この係数を考慮せず学習したものと 1000 局対戦した勝率を図 4 に示す。

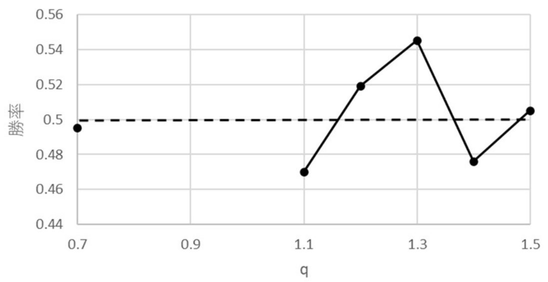

図4 パラメータqと勝率

係数ｑは１未満のほうが勝率は良くなると考えていたが、1 より大きい条件で有利なものがあった。ただし、不安定な結果となった。

追試可能か

学習は、教師局面をランダムにシャッフルしているため、全くの再現は難しいと考える。

また、本アイデアの効果は、学習モデルの違い、学習の成熟度、学習データ、各種パラメータの設定など様々な要素により異なると思われ、同様の結果が再現するかはよくわからない。ただ、現局面ではなく、静止探索の末端局面に対して学習させると強くなる理由は解明されておらず、今後も研究対象となると思われる。

以上