第30回世界コンピュータ将棋選手権

「SNNS」アピール文書

2020年月日

芝浦工業大学Ⅱ

吉野　拓真

SNNSは、モンテカルロソフトマックス探索とNNによる評価関数表現を組み合わせた構成となっています。プログラムは、山岡忠夫さんの「将棋AIで学ぶディープラーニング」[1]のモンテカルロ木探索とNNを組み合わせたものをベースとしています。以下、特徴・工夫点について記述します。

1. モンテカルロソフトマックス探索[2]を使用

　元のプログラムでは、探索部にAlphaGoZero同様のモンテカルロ木探索に基づいた探索を用いています。本プログラムでは、モンテカルロソフトマックス探索を用いています。モンテカルロソフトマックス探索は、選択探索の一種で、以下の4ステップを繰り返します。

* + 1. 選択
		2. 展開
		3. 評価
		4. バックアップ

　モンテカルロソフトマックス探索は、「ノードを確率的に選択する」、「子ノードの展開を先に行い、それらのノードをまとめて評価する」等の点で、モンテカルロ木探索とは異なる性質をもちます。

1. 選択評価値

　モンテカルロソフトマックス探索にてPolicy networkの出力をノード選択に活かすため、以下の$E\_{s}\left(v\left(a;s\right)\right)$を「局面ノードsにおける合法手aを指した後の局面子ノード$v\left(a;s\right)$に対する評価値」としてノード選択にのみ導入し、これを選択評価値と呼びます。

 $E\_{s}\left(v\left(a;s\right)\right)=E\left(v\left(a;s\right)\right)+c\_{puct}P\left(s\right)\sqrt{\frac{\sum\_{b\in A(s)}^{}n\left(b|s\right)}{1+n\left(a|s\right)}}$

　$E\left(v\left(a;s\right)\right)$は子ノード$v\left(a;s\right)$の評価値、$c\_{puct}$は定数、$P\left(s\right)$は局面ノードsにおける合法手aに対するPolicy networkの出力値、$n\left(a|s\right)$は局面ノードｓで指し手aを選択した回数とします。式の第2項は、AlphaGoZeroのPUCT式の第2項と同様のものとなります。

　ノード選択における選択確率$π\_{s}\left(a|s\right)$の計算にこの選択評価値$E\_{s}\left(v\left(a;s\right)\right)$を用いて、

$$π\_{s}\left(a|s\right)=exp\left(\frac{E\_{s}\left(v\left(a;s\right)\right)}{T\_{s}}\right)/Z\_{s}$$

$$Z\_{s}≡\sum\_{x\in A\left(s\right)}^{}exp\left(\frac{E\_{s}\left(v\left(x;s\right)\right)}{T\_{s}}\right)$$

とします。$T\_{s}$は、選択確率の計算に用いる温度パラメータです。

1. 評価

評価関数は、著書のものをそのまま使用しています。ノード展開時に新たに生成された各子ノードの入力特徴を作成し、それらをまとめて評価します。モンテカルロソフトマックス探索では、1回のシミュレーションで複数のノードをまとめて評価するため、GPUによる並列計算と相性が良いと言えます。

　元のプログラムでは1シミュレーション毎に1つのノードに対して入力特徴の作成を行いますが、本プログラムでは各子ノードに対して行うので、親ノードとの差分を用いる事により入力特徴作成の効率化を図っています。。

[1]山岡忠夫，“将棋AIで学ぶディープラーニング”，マイナビ出版，2018

[2] 桐井杏樹，原悠一，五十嵐治一，森岡祐一，山本一将，“確率的選択探索の将棋への適 用”，第 22 回ゲーム・プログラミング・ワークショップ 2017 予稿集，pp.26-33，2017.