GLOBIS-AQZ アピール文書

GLOBIS-AQZは株式会社グロービス、株式会社トリプルアイズ、山口祐の三者が開発を進める囲碁AIです。2018年9月に開発プロジェクトを発足、囲碁AI世界一と若手棋士育成を目指しています。

開発にあたってはマルチエージェントによるShared Population Based Training、最大1120 GPUによる大規模分散強化学習、中国ルールから日本ルールへの転移学習を行っています。

1. マルチエージェント強化学習

AlphaGo Zero[1]に代表される深層強化学習は、ドメイン知識を用いることなく優れた戦略を獲得することができます。しかし学習が進むに連れ、エージェントの戦略が固定化されたり、三すくみの戦略を循環するなど、局所解に陥る可能性が高くなります。Population Based Training（PBT）は複数のエージェントを並行的に学習させ、ハイパーパラメータを遺伝的アルゴリズムにより最適化することで、シングルエージェントの場合より効率的に学習させることができる手法です[2,3]。一方でエージェントごとに学習が必要なので、計算コストはエージェントの数だけ増大するという問題もあります。

GLOBIS-AQZはPBTを発展させた「Shared Population Based Training」を提案し、囲碁の強化学習に応用しています。PBTと同様、複数エージェントでハイパーパラメータを遺伝的に改善しながら学習させますが、生成した学習データをエージェント間で共有し、計算量を抑える手法です。自分と他エージェントが生成したデータの学習率比もハイパーパラメータとして保持しており、他のエージェントの戦略を最適な割合で取り入れています。

2. 大規模分散強化学習

GLOBIS-AQZは産業技術総合研究所のAI橋渡しクラウド（ABCI）を利用し、最大1120GPUを用いた強化学習を行っています。100GPUを超える大規模の並列強化学習を行う場合、推論やパラメータ学習だけでなく、生成した学習データを管理する部分もボトルネックとなりえます。

ABCI上に自己対戦サーバ（TeslaV100最大1024基）、学習サーバ（最大32基）、レーティング計測サーバ（最大64基）を配置し、これらを集中的に管理するデータベースシステムを用いてスループットを最大化するように設計しています。これにより、毎秒2500以上生成される局面データに対しても連続的に強化学習することを可能にしています。

3. 日本ルールへの転移学習

一般に、コンピュータ囲碁では日本ルールより中国ルールの方が容易に終局判定ができます。地を数える日本ルールでは、不要な手入れをせずに適切にパスをする必要があるためです。これはニューラルネットワークを用いた深層強化学習でも同様で、半目勝負の終局時のみ有効手となるパスを適切に学習するのは簡単ではありません。

GLOBIS-AQZはニューラルネットワーク以外にも古典的なモンテカルロシミュレーションも併用しており、日本ルールでも正確に終局判定を行うことができます。中国ルールコミ7目半で学習したニューラルネットワークに、終局判定のみを変更した強化学習を追加で実施することで、日本ルールコミ6目半でも正確な盤面評価を行えるように学習を進めています。

[1] D. Silver et al. (2017) Nature, 550, pp.354-359
[2] M. Jaderberg et al. (2017) https://arxiv.org/abs/1711.09846
[3] M. Jaderberg et al. (2019) Science, 364, 6443, pp.859-865