

論文の内容の要旨

論文題目 モンテカルロ木探索の 改善に関する研究

氏名 今川 孝久

本論文は二人零和完全情報ゲームを主な対象とした、モンテカルロ木探索 (MCTS) の研究について記したものである。

ゲームにおいては、全ての局面を読み切ることが一般的である。それは、一般に、ゲームは探索空間が広く、かつ手を打つまでの時間が限られているためである。従って、探索では、効率的に最善手を見つけることが要請される。MCTS は、モンテカルロシミュレーションを繰り返し、その結果が良い選択枝を優先的に探索するアルゴリズムである。MCTS では、シミュレーション結果に対し、利得 (良さ) を与え、各候補手を利得の平均値で評価し、評価の高い手を優先して、先読みとシミュレーションを行う。

最善手の判別のためには、各候補手を選んだ場合の真の利得の期待値を推定する必要がある。代表的な MCTS では、多腕バンディット問題 (MAB) における、累積リグレット (累積的な損失) の最小化のためのアルゴリズムを、シミュレーションを開始する節点の決定に応用している。MCTS に累積リグレットの最小化のためのアルゴリズムを応用することで、MAB に同様に、徐々に最善手を多く選ぶようになり、シミュレーションの平均的な利得が、最善を尽くした場合の利得に近づくこと期待される。

MCTS は汎用なアルゴリズムであり、General Game Playing や、General Video Game Playing という初見のゲームを総合的に上手くプレイすることが求められるドメインで成果を上げてきた。また、囲碁や Hex といった個々のゲームでも主流な探索手法である。

特に囲碁では、初めてプロ棋士に勝ったプログラム AlphaGo での探索に採用され、その勝利に貢献した。

しかしながら、MCTS は常に優れた性能を発揮出来る訳では無く、比較的苦手なドメインがあることが知られている。そして、MCTS の性能とドメインの特徴の関係はまだ不明な点も多く、

その関係性を明らかにする必要がある。また、代表的な MCTS ではシミュレーションを開始する節点を選ぶことを MAB とみなして、MAB のアルゴリズムを用いているが、MAB と木探索では異なる点があり、MAB の性質は成り立たない。そのため、MAB より木探索の性質に近い仮定とその仮定に基づくアルゴリズムについて調べる価値がある。

本研究では、MCTS に関して、ドメインの種類とそこでの性能、理論的な仮定といった観点から議論し、既存手法の問題点を指摘し、その改善策を提示した。具体的には、四つの点に着目して研究を行った。

第一点目として、MCTS が比較的苦手とされるドメインの特徴の一つとして、最善手を選び損ねた場合の不利益の偏りという特徴に新たに着目する。そして、その特徴と MCTS の性能の関係性について分析する。MCTS が成果を上げてきたドメインの一つである、囲碁の中でも、攻め合いの局面は比較的苦手であると知られている。本研究で扱う特徴を持つ局面は、攻め合いの局面と共通の難しさを有している。本研究では、シミュレーションの結果、囲碁での目数の差といったゲーム上のスコアが手に入るという仮定の下で、理論的な分析に基づく利得の与え方を提案する。そして、既存手法が性能を発揮しづらい特徴を持つ局面で、提案手法が効果的であることを示した。

第二点目として、MCTS における探索資源の割り振りに着目し、直接的に単純リグレット(最終的な手の選択での損失)の最小化を目指した MCTS を提案する。探索の目的は最善手を判別することであり、これは単純リグレットを最小化することである。しかしながら、探索において、MAB のアルゴリズムを使い、単純リグレットを直接最小化することは難しいため、累積リグレットの最小化のための手法が MCTS に応用されてきた。本研究では、探索での単純リグレットの直接的な最小化に向け、分布を使った MCTS を提案する。提案手法は、扱う分布を真の利得の主観確率の分布と仮定すると、最善手の判別のために、探索資源の割り振りが効果的になされると期待できる手法である。実験の結果、提案手法の性能は、深さが一様な木では、既存手法を超えないものの、非一様な木では上回った。

第三点目として、MCTS における勝敗の確定に着目する。探索木中のある節点が終端(ゲームの終わり)の場合、その節点での勝敗が確定する。これは、シミュレーションによる確率的な評価しか得られない MCTS において、例外的に確実な評価が得られるという意味で重要である。しかしながら、既存手法では、その情報を枝刈りにしか利用していないという問題点がある。本研究では、勝敗確定の情報を枝刈りだけでなく、推定値の修正とそれに応じた探索の優先度度合いの調整に利用する方法を提案する。計算機実験により、提案手法は既存手法よりも効果的であることを示した。

第四点目として、MCTS における候補手の利得の推定値に着目する。最善手の判別のためには、最善を尽くした時の利得を正確に推定する必要がある。特に、最善手は以後最善を尽くすと仮定した場合での最も良い利得の手であるので、期待利得の最大値の推定が重要である。まず、MAB の設定下で、利得の期待値の最大値に対する、新しい推定量 **Simplified Weighted Estimator (SWE)** を提案する。各選択枝の試行回数が予め決まっているという仮定

のもとで, SWE について理論的な分析を行い, 試行回数が増えるにつれて, 推定値が正しい値に収束することを示した. さらに, 累積リグレット最小化のためのアルゴリズムに基づいて試行回数が決まる場合での, 推定値の正確さについて, 実験を行い検証した. その結果, 設定次第では, 既存手法の性能が SWE を超える場合もあるが, 設定に対する推定量の性能の安定性という観点で SWE が優れていることを示した. 特に, MCTS で行われている, 平均による推定値は, 正しい値と比べて過度に低くなることを示した. 加えて, SWE を MCTS に応用し, 実験で性能を確かめる. 実験は, 終端節点で利得が確率的に得られる木と確定的に得られる木の2種類で行った. 提案手法は後者の木では, 既存手法を超えないものの, 前者の木では上回る性能を示した.

以上のように本研究は, 複合的なアルゴリズムである MCTS について, 利得の与え方, 候補手の評価の仕方, 評価に対する探索資源の割り振り方という三つの要素技術のそれぞれの観点から議論し, 改善策を提示したと言える.