

Title	ターン制戦略ゲームへの深層学習の適用
Author(s)	木村, 富宏
Citation	
Issue Date	2020-12
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17049
Rights	
Description	Supervisor:池田 心, 情報科学研究科, 博士

氏 名	木村 富宏		
学 位 の 種 類	博士(情報科学)		
学 位 記 番 号	博情第 442 号		
学 位 授 与 年 月 日	令和 2 年 12 月 24 日		
論 文 題 目	ターン制戦略ゲームへの深層学習の適用		
論 文 審 査 委 員	主査 池田 心	北陸先端科学技術大学院大学	准教授
	飯田 弘之	同	教授
	長谷川 忍	同	准教授
	白井 清昭	同	准教授
	鶴岡 慶雅	東京大学	教授

論文の内容の要旨

This paper applies a new approach to analyze game research, deep learning, and turn-based strategy games through the creation of artificial intelligence(AI) capable of strong play. It identifies what is needed to achieve this, why older algorithms have failed to perform, and what technical issues need to be addressed and solved for algorithms to perform using deep learning. AI is an important discipline that has been studied for many years and has a wide range of applications. Games are widely used in AI research because of being relatively easy to reproduce and evaluate, for which techniques such as tree search, Monte Carlo methods, and machine learning have been applied or proposed. In recent years, breakthroughs such as AlphaGo and AlphaZero have made AI research more active.

Turn-based strategy games are one of the most popular fields in the world and are more complex and challenging than Go and shogi. In particular, since each player can move all of their pieces on the board in any order, the number of player choices increases exponentially, resulting in the so-called computational explosion, which is difficult to solve in AI design. In addition, turn-based strategy games are difficult to analyze because of the complexity of the data, which is multilayered due to the nature of the pieces and terrain. For these reasons, to date there has been limited progress in AI research for turn-based strategy games. In this paper, an algorithm implementing deep learning proposed by AlphaZero is applied to solve the above difficulties and attempts to advance the research on AI in turn-based strategy games.

In this thesis, the existing research on games was surveyed, including Monte Carlo tree search (MCTS), which is a commonly used algorithm in game research, realized probability search, and board evaluation function using machine learning. Furthermore, approaches such as AlphaGo/AlphaGoZero/AlphaZero, which combine deep learning and reinforcement learning, are also discussed.

In order to perform the study in TUBSTAP efficiently, we first prepared a number of benchmark maps. A benchmark map is a map of relatively simple problems from a human perspective, which we used to check the basic performance of the algorithms we created. They were designed and produced for TUBSTAP and contained a total of 8 groups of 43 maps, which we presented at the 2016 Game Programming Workshop. We investigated the performance and limitations of the existing algorithms using the benchmarking problems and found that they often fail to solve even the problems that can be easily solved by humans. This included programs which had won at the highest level of competition. Expressing complex data

structures and behavioral expressions in turn-based strategy games using deep learning neural networks is a challenge in actual practice. In this research, we introduced a recurrent network that can output data in a time-division manner to concisely express a complicated and nested data structure and reduce the number of neural network outputs. To evaluate this network, we tried to create a computer player by reinforcement learning. The profit sharing method, which shows a sharp rise to sparse rewards, was selected as the reinforcement learning method, and as a result of learning about 20 or more maps for the passfind and tracking maps examined in the benchmark problem, the episode success rate was 20%-30%. This was higher performance than the DQN algorithm.

Next, the issues in applying the deep learning algorithm proposed by AlphaZero to turn-based strategy games were considered and systematically investigated. This algorithm is referred to as PV-MCTS in this paper. To implement PV-MCTS, the data structure of the neural network is moved from the output to the input side of the neural network for unit selection to reduce the output layer, and the search is accelerated by adding an attack bias term to the polynomial upper confidence tree (PUCT) value. The neural network is not a recurrent network but is designed with eight layers of Residual blocks that were used in AlphaZero. In this design, multiple layers are used to enhance the expression in complex situations. Performance and competitive evaluation of various parameters of the PV-MCTS algorithm were carried out. In the battle evaluations, we won more than 80% of the matches against our original MCTS algorithm, and also showed a high win rate on unknown maps.

Key words: Artificial Intelligence, turn-based strategy game, deep learning, AlphaZero, Monte Carlo tree search

論文審査の結果の要旨

ゲームはルールが明確に定義され、勝敗を通じて強さの評価が容易であるなど、人工知能技術の良い適用対象である。機械学習・最適化・木探索など多くの手法がゲームを通じて開発・評価・改善されてきた。深層学習は近年最も注目を集める技術領域であり、なかでも AlphaZero アルゴリズムは自己対戦学習のみで囲碁・将棋・チェスのトップ棋士や既存プログラムに勝てることが示されたことで代表的な手法となっている。一方で、ゲームには多種多様なものがあり、AlphaZero アルゴリズムなど既存の手法がただちには適切に適用できないような領域も存在する。ターン制戦略ゲーム(TBS)は複数の駒を任意の順序で動かせるなどいくつかの特徴を持ち、適用困難なゲームジャンルの一つである。

このような背景のもと木村さんはまず TBS というジャンルが囲碁や将棋に比べ困難である理由を分析し、TBS の典型的なルールのみを採用した TUBSTAP という学術用プラットフォームを採用した。そのうえで、封鎖・逃走・経路探索・護衛・挟み撃ち・待ち伏せなどの TBS の基本的な戦術を盛り込んだ「ベンチマークマップ」を多数生成、公開のうえ、既存プログラムの挙動を評価した。これにより、既存プログラムの限界と課題を明らかにし、それに多くの研究者が取り組めるようにした。

分析の結果、最も大きな課題は「行動が移動元・移動先・攻撃先と階層的で複雑であること」「ランダム性の高い行動では報酬が稀にしか得られないこと」であると分かった。木村さんはこ

の課題に対しまず第 5 章では, RNN を用いて階層的な出力構造を効率良く表現すること, Profit Sharing の考え方を用いて報酬が得られた稀なエピソードを最大限活用すること, で標準的な CNN を用いた DQN 強化学習よりも性能が高くなることを示した. 続いて第 6 章では RNN による表現が強化学習のみならず policy network の教師付き学習に用いても高い性能を示すことを示した. 学習された policy network は, 既存の木探索プログラムに対し, 探索なしでも互角以上の勝率となった.

第 7 章ではこれまでの知見を活かし, AlphaZero アルゴリズムに適切な工夫を加えることで TBS への適用を果たした. 1) 移動駒選択を出力ではなくネットワークの入力側に移して設計を簡略化する, 2) 手筋マップを学習環境に加えることで報酬を得るための支援を行う, 3) 攻撃行動に探索時の選択バイアスをかけることで探索を高速化する, などの工夫を加えることで, 既存アルゴリズムに対し同じ探索時間の対戦で大きく勝ち越し, 未学習マップにおいても汎化能力によって互角以上に戦えることを示した.

以上, 本論文は複雑なゲームへの深層学習の適用可能範囲を広げたものであり, 学術的に貢献するところが多い. よって博士 (情報科学) の学位論文として十分価値あるものと認めた.