

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	<a href="http://hdl.handle.net/10119/17049">http://hdl.handle.net/10119/17049</a>
Rights	
Description	Supervisor:池田 心, 情報科学研究科, 博士

## Abstract

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