

Mastering the game of Abalone using deep reinforcement-learning and self-play

Ture Claußen

Bachelor thesis in "Applied computer science"

November 29, 2021



Author Ture Claußen
Matriculation number: 1531067
tu.cl@pm.me

First examiner: Prof. Dr. Adrian Pigors
Abteilung Informatik, Fakultät IV
Hochschule Hannover
adrian.pigors@hs-hannover.de

Second examiner: Prof. Dr. Vorname Name
Abteilung Informatik, Fakultät IV
Hochschule Hannover
email-Adresse

Declaration of authorship

I hereby declare that I have written this thesis independently without any help from others and without the use of documents or aids other than those stated. I have mentioned all used sources and cited them correctly according to established academic citation rules.

Hannover, November 29, 2021

Signature

Contents

- 1 Introduction 8**
 - 1.1 Research goals 9
- 2 Analysis 10**
 - 2.1 Artificial intelligence 10
 - 2.1.1 Rational agent 10
 - 2.2 Environment 11
 - 2.2.1 Abalone rules 11
 - 2.2.2 Abalone complexity 12
- 3 System architecture 14**
 - 3.1 Software 14
 - 3.1.1 Training framework 14

List of Figures

2.1	The agent-environment interaction loop	11
2.2	Basic moves [S.A]	11
2.3	Sumito positions allow pushing the opponent's marbles [S.A]	12
2.4	Counts of moves available for random for random player in 5 games . . .	13

List of Tables

2.1 Abalone in comparison with other games [Cho09] 13

Abstract

Explanation

1 Introduction

Board games are and have been a popular environment to test the capabilities of state of the art artificial intelligence against human opponents. Many board games are widely known making them a tangible measure of performance. The most prominent examples are the games of Chess and Go. For both, machines defeating the current best players has been representative of fundamental progress in computing.

IBM's "Deep Blue" defeated Gary Kasparov in 1996 [Hig17] by utilizing search to look ahead into the game tree and deliberate on the next move. This approach is a prime example for symbolic AI approaches, "good-old-fashioned-AI" ("GOFAI") [Hau85], which rely on logic and search on symbolic representations.

However, these knowledge-based approaches are severely limited by our ability to properly model the problem correctly and exhaustively. For example, in the case of Deep Blue it requires us to encode our knowledge about chess in a heuristic function to evaluate the board. Only then we can search for actions that maximize this function. Problems with large complexity would require tremendous efforts, which just become unfeasible at a certain point. A different approach would be devising (general) methods to learn the necessary domain knowledge from scratch, *tabula rasa*. As Alan Turing put it:

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child-brain is something like a note-book as one buys it from the stationers. Rather little mechanism, and lots of blank sheets. [...] Our hope is that there is so little mechanism in the child-brain that something like it can be easily programmed. [TUR50]

The recent success of "AlphaGo" in 2016 against the long-time world-champion Lee Sedol [Dee] in the game Go is a milestones that perfectly demonstrates this shift towards "bottom-up" or subsymbolic methods. [Nil98] The increasing availability in computational power (and data) has enabled two subsymbolic methods to find large success in unclaimed territory such as computer vision or natural language processing. Namely those are neural networks and (stochastic) gradient descent. Combined they provide a general function approximator, that can be trained in a process akin to the learning described by Turing.

In the case of Go, designing a powerful heuristic function was deemed not possible for humans. AlphaGo used (deep) neural networks and gradient descent to train a evaluation function based on a large database of expert moves. With the help of Monte Carlo Tree Search they used this function to play against itself and improve further. [SSS⁺17]

Building on this success DeepMind, the company behind AlphaGo, further improved the architecture. "AlphaGo Zero" and the generalization "AlphaZero" learn, without the help of the database of expert moves and surpassed the performance of AlphaGo significantly. Since then the architecture has been applied to Chess, Shogi and Atari games by removing the last piece of human knowledge in the system: The rules of the game. [SAH⁺20]

At this point our endeavor begins, as the purpose of this writing is to apply the methods of AlphaZero to the game of Abalone.

1.1 Research goals

First, let us establish the main research questions that will guide us throughout this thesis.

The goal is to apply the general framework of self-play learning outlined in "Mastering the game of Go without human knowledge" to the board game of abalone. [SSS⁺17] The original paper gives clear instructions on the theoretical groundwork for the system but omits clear instructions for the implementation. There is no open source code provided.

Sub-goal 1 is to compare classical search based methods to this AlphaZero's deep reinforcement learning based on several criteria such as win/loss ratio, computational requirements, etc.

2 Analysis

Before we move to the nuts and bolts of AlphaZero and our concrete implementation, we should establish a general understanding of the problem. That includes establishing the necessary theoretical background in artificial intelligence in general, as well as insight into the specialized knowledge such as deep reinforcement learning.

2.1 Artificial intelligence

2.1.1 Rational agent

Stemming from the latin word *agere* meaning "to act", an agent is something that acts. As we expect our agent to take sensible or intelligent actions we further qualify this definition by calling it rational. This means that it acts so as

to achieve the best outcome or, when there is uncertainty, the best expected outcome.

[RN21, p. 36]

The agent exists in an environment which it perceives through sensors and it acts takes actions through its actuators. We refer to the content of the sensors output for one observation as *percept*. The cat uses eyes, ears and other sensors to perceive the world and its legs, claws and so on to interact with the world. This forms a loop of observing, deliberating and finally taking an action, whereafter the agent observes again as depicted in figure 2.1. Internally our agent has some an *agent function* that takes the entire history of percepts observed and the built-in knowledge and maps it an action. A concrete implementation of this abstract function is called *agent program*.

The agent program might just be a simple tabular mapping from percepts to actions or could use a complex model of rules about the environment.

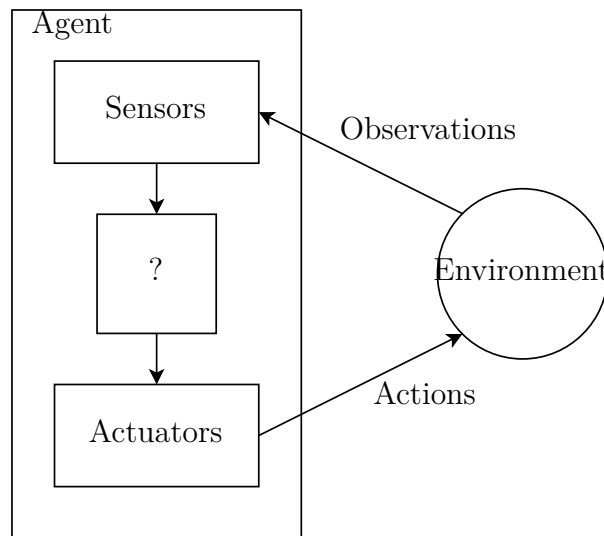


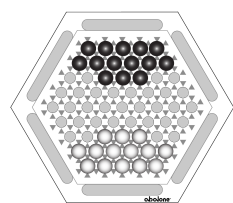
Figure 2.1: The agent-environment interaction loop

2.2 Environment

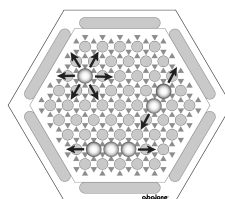
2.2.1 Abalone rules

Abalone is a fairly new game, that was devised in 1987 by Michel Laet and Laurent Lévi. Nevertheless, with more than four million global sales it has established itself as a classic game [noa20]. Abalone is a two-player game consisting of a hexagonal board with 61 fields and 14 marbles for black and white respectively.

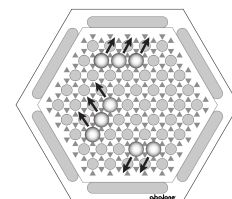
The goal of the game is to push six of the opponent's marbles off the playing field. The game's starting position is depicted in figure 2.2 (a). One, two, or three adjacent marbles (of the player's own color) may be moved in any of the six possible directions during a player's turn. We differentiate between broadside or "side-step" moves and "in-line" moves, depending on how the chain of marbles moves relative to its direction, which is shown in figure 2.2 (b) and (c).



(a) Starting position



(b) "In-line" moves



(c) "Side-step" moves

Figure 2.2: Basic moves [S.A]

A move pushing the opponent's marbles is called "sumito" and comes in three variations, as shown by figure 2.3. Essentially, the player has to push with superior numbers and the opponent's marbles can not be blocked. This is the game mechanic that allows for pushing the marbles out of the game and winning.

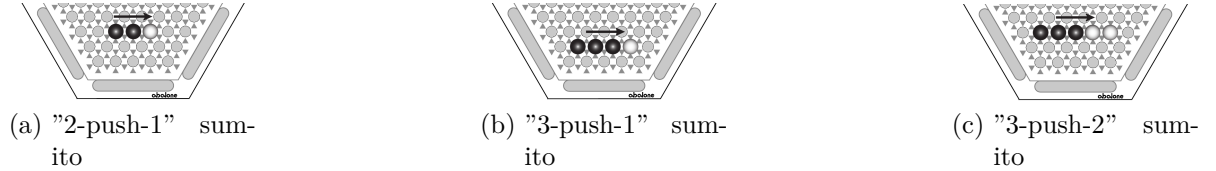


Figure 2.3: Sumito positions allow pushing the opponent's marbles [S.A]

2.2.2 Abalone complexity

An important characteristic of a game environment is its complexity, which can be described in two relevant dimensions.

State space complexity The state space is the set of all possible states the agent can be in.[RN21, p. 150] For Abalone this means we have to consider all possible board configurations with different numbers of marbles present. Additionally, we would have to remove duplicates that arise from the symmetries of the board. In the case of abalone we have 6 rotations and 6 axes we can mirror the board on. The following formula gives us a good upper bound:

$$\sum_{k=8}^{14} \sum_{m=9}^{14} \frac{61!}{k!(61-k)!} \times \frac{(61-k)!}{m!((61-k)-m)!}$$

Game tree complexity The game tree defines the dependencies between board positions (nodes), moves (edges) and the resulting board positions. The number of nodes in a tree is given by:

$$b^d$$

First we consider the branching factor b , or the number of possible moves for any given state. We can only approximate this, as this number greatly varies between different states. On average abalone has $b = 60$ possible moves per state as measured in figure 2.4. The depth d of the tree depends the number of turns per game. For some leaves this might be quite early, but as Abalone allows for actions to be taken repeatedly the

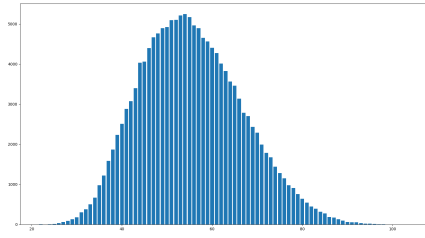


Figure 2.4: Counts of moves available for random for random player in 5 games

Game	state-space complexity (log)	game-tree complexity (log)
Tic-tac-toe	3	5
Reversi	28	58
Chess	46	123
Abalone	24	154
Go	172	360

Table 2.1: Abalone in comparison with other games [Cho09]

height is potentially infinite. Looking at the average again a game takes in the region of $d = 87$ turns, giving us a total of 60^{87} nodes. [Lem05]

As those numbers in isolation are hard to grasp it is useful to put Abalone's complexity in relation with other popular games. Its state space complexity is on the same level as Reversi, whilst its game tree surpasses chess in complexity (c.f. table 2.1)

3 System architecture

3.1 Software

3.1.1 Training framework

As there are existing frameworks that have implemented the system described in the AlphaZero paper in a more general and adaptable fashion, it has to be considered building on their foundation.

Bibliography

- [Cho09] Pascal Chorus. Implementing a computer player for abalone using alpha-beta and monte-carlo search. Master’s thesis, Citeseer, 2009.
- [Dee] DeepMind. Match 1 - Google DeepMind Challenge Match: Lee Sedol vs AlphaGo. <https://www.youtube.com/watch?v=vFr3K2DORc8&t=7020s>.
- [Hau85] John Haugeland. *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, Mass, 1985.
- [Hig17] Chris Higgins. A Brief History of Deep Blue, IBM’s Chess Computer — Mental Floss. <https://web.archive.org/web/20170803130439/https://www.mentalfloss.com/article/503178/brief-history-deep-blue-ibms-chess-computer>, July 2017.
- [Lem05] NPPM Lemmens. Constructing an abalone game-playing agent. In *Bachelor Conference Knowledge Engineering, Universiteit Maastricht*. Citeseer, 2005.
- [Nil98] Nils J. Nilsson. *Artificial Intelligence: A New Synthesis*. Elsevier, April 1998.
- [noa20] Abalone (board game). [https://en.wikipedia.org/w/index.php?title=Abalone_\(board_game\)&oldid=987654321](https://en.wikipedia.org/w/index.php?title=Abalone_(board_game)&oldid=987654321), December 2020.
- [RN21] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc, fourth edition, 2021.
- [S.A] Abalone S.A. Abalone rulebook. <https://cdn.1j1ju.com/medias/c2/b0/3a-abalone-rulebook.pdf>.
- [SAH⁺20] Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, 588(7839):604–609, December 2020.
- [SSS⁺17] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, October 2017.

- [TUR50] A. M. TURING. I.—COMPUTING MACHINERY AND INTELLIGENCE.
Mind, LIX(236):433–460, October 1950.