



ÉCOLE
POLYTECHNIQUE
DE BRUXELLES

Techniques of artificial intelligence

PROJ-H418

**Project report : *Monte-Carlo* tree-search for
Checkers**

Sami ABDUL SATER
Alexandre FLACHS
Diego RUBAS
Jeanne SZPIRER

Academic year 2021-2022

Contents

1	Introduction : <i>Monte-Carlo</i> tree-search	1
1.1	Parameters	2
1.2	Optimization and constraints	2
1.3	Our contribution	2
2	Rules of the Checkers game	2
2.1	Beginning of the game	3
2.2	Movements	3
2.3	Endgame	3
3	Implementation of MCTS to Checkers	3
3.1	Nodes	3
3.2	Selection policy	4
3.3	Expansion	4
3.4	Best child policy	4
3.5	Simulation : introducing heuristics	4
3.5.1	Heuristic 1	5
3.5.2	Heuristic 2	5
3.5.3	Heuristic 3	5
3.5.4	Heuristic 666	5
4	Genetic Algorithm to tune parameters	5
4.1	Parameters to find	5
4.2	Algorithm	5
4.2.1	First generation of parents	5
4.2.2	Simulation of games	5
4.2.3	Evolution of the population	5
4.2.4	Continuation and end	6
5	Benchmark and analysis of the AI	6
5.1	Initial parameters and approach	6
5.2	Benchmark setup and implementation	6
5.3	Wins and losses	6
5.4	Execution time behaviour depending on the parameters	7
A	Listings	7

1 Introduction : *Monte-Carlo* tree-search

Tree search is an intuitive way to solve a game with a limited number of possible moves. A *Monte-Carlo* tree-search (MCTS) is a tree-search algorithm that exploits **randomness** and **evaluation of simulated games** to decide the next move. The tree is built according to a policy that we hereby define.

Repeat n_{iter} times :

1. **Selection** of the **best** node according to policy
 - **Expansion** of nodes if needed
2. **Simulation** of the rest of the game, starting from the selected node. This simulation ends with a **reward** that takes into account if the game has been won or not.

3. This reward is **backpropagated** to the selected node.

Once all the simulations have been done, the tree is considered to be computed (though not necessarily fully expanded) : we then select the **best child**.

1.1 Parameters

Are variable :

- The selection policy
- The best-child selection policy
- The number of iterations

1.2 Optimization and constraints

There are no particular mathematical constraints to ensure for this project. However, constraints are to be imposed to make it sure it runs in a **realistic time**, e.g. 15 seconds by move.

Under this time, the parameters of the search (n_{iter} , the policies, and more) must be tuned to **optimize the win rate**.

This report presents the implementation of a MCTS on top of a Checkers game. Explaining first the rules, very briefly, we then explain the implementation itself before presenting results of our AI against a **deterministic** AI (minimax).

1.3 Our contribution

We took the implementation of a Checkers game with a minimax AI on top of it from an Open Source repository. Implementing MCTS required a huge refactor, at the game level and thus also at the minimax level. After implementing MCTS and refactoring, a benchmark was run for different parameters, which lead to an optimization of the win rate over the parameters of the search.

2 Rules of the Checkers game

Let's briefly go through the rules of Checkers game. Particular terms will be used and highlighted, that will be important for the algorithm.

The game opposes two adversary, here named **RED** and **WHITE**, and consists of a **board** and **pieces** on it, each belonging to one player. Each piece then has a color, and the board has pieces on it. Here are some additional information :

- A board can call a function to get all the pieces of a certain color
- It is possible to move a piece of the board using a **move** function
- A piece has a defined **position** (x, y) where x denotes the row and y the column of the piece. A piece is hence defined by a color and its position : $P = (C, x, y)$

2.1 Beginning of the game

Each player has 12 pieces, that begin at the same position for every game, and the starting position is the conventional position for Checkers game. From here, the **WHITE** begins (by convention) and can perform a move.

2.2 Movements

In this section, we define with words how a player can move a piece. We could define it in an algorithmic way, but this wouldn't be particularly relevant for the sake of this report.

A player can only **move** a piece in diagonal, going forward, and can move only one row forward, unless an ennemy piece is on its way. In this case, if the piece can reach a place on the board and some enemy pieces are on its way, the enemy pieces are discarded and the initial piece can find its final destination. We say that the pieces has **skipped** n pieces if n enemy pieces were discarded.

If a piece reaches the opposite side of the board, it becomes a **queen** and can from now on move backwards.

2.3 Endgame

A game ends when

- a player has no pieces left : the adversary wins ;
- all remaining pieces are queens and no piece was discarded in the last 20 moves : it ends as a draw.

3 Implementation of MCTS to Checkers

To implement the tree search, we need to define a tree and the policies associated to the search. Each time the AI has to play, it calls the algorithm, beginning to build the tree (as described in the introduction). Once the tree is built, it selects the best move according to a policy that will be described further.

3.1 Nodes

A node in the tree corresponds to

- The parent node
- A **state** of the game : a **board** element
- The move that lead from previous node to this one (*parent action*) : a **move** element
- **visits** : number of times that this node was visited during the search
- **reward** : number of times that this node led to victory

When the AI is instanciated, the root node has no parent and no parent action, **visits** is set to 1 and **reward** is set to 0.

The resulting constructor for the class **MCNode** can be found on listing 1.

3.2 Selection policy

To select a child node from which perform a simulation, the current node first needs to check if it has children, and if so, if all children have been explored. That is, the current node builds a list of possible children (resulting from the possible moves) and looks for children that are not currently in the tree.

Thus, if the node is currently not **fully explored**, we **expand** the current node. If the node is fully explored, we select the **best child** to perform the simulation.

Intuitively, if there is a sequence of fully explored nodes that leads to a leaf, this will be the privileged path. Otherwise, the algorithm will return the first created child node of a non-fully-explored node. This yields in listing 1.

3.3 Expansion

When a node N needs to be extended, first a list of possible moves is created. Then, a **random move r is drawn**¹ and a node N_r is created from this move, having N as **parent node** and r as **parent action**. This yields in listing 2.

3.4 Best child policy

There are multiple calls to the best-child policy :

- When the tree is built and we need to perform an actual choice : we choose the best child node of the root node
- During the selection, when a node is fully explored and we need to go down a level to look for a leaf node to select or a node to expand.

During the exploration of the tree, the attributes **reward** and **visits** of each node are updated, such that at any time, it is possible to define, for each node, a value evaluating the node. We chose the following values² :

$$\begin{aligned} \blacktriangleright \text{Exploration } d(N) &= \frac{N.\text{reward}}{N.\text{visits}} \\ \blacktriangleright \text{Exploitation } e(N) &= \sqrt{\frac{\log_2 N.\text{visits}}{N.\text{visits}}} , \\ \blacktriangleright \text{Score } s(N, w_e) &= d(N) + e(N) \cdot w_e \end{aligned}$$

where w_e is a parameter to define, to make the famous trade-off between exploitation and exploration.

Once this score computed for each child node, we select the one with the best score (or a random one among the equivalent children).

3.5 Simulation : introducing heuristics

Once a node is selected, we can simulate a game until a final position is reached. The game simulation is done by playing heuristic-based moves. We defined some heuristics to evaluate

¹pour l'instant. a modifier si on change

²Insérer source wikipédia

the quality of a move, and the possible moves are sorted according to this evaluation.

Hence, the game can be simulated by choosing the best move, generating a new board, and changing the player's turn accordingly.

3.5.1 Heuristic 1

3.5.2 Heuristic 2

3.5.3 Heuristic 3

3.5.4 Heuristic 666

4 Genetic Algorithm to tune parameters

4.1 Parameters to find

The implemented genetic algorithm allows to calculate the value of some parameters in order to improve the main Monte-Carlo algorithm which plays checkers. These parameters are :

- the number of iterations
- the safety factor (see Safe heuristic)
- the exploitation factor.

Except for the number of iterations, the factors to be found can be between 0 and 1 and are there to determine the importance of the coded heuristics.

4.2 Algorithm

4.2.1 First generation of parents

The algorithm starts with a random generation of "Villagers" (representing the population). *POP_SIZE* villagers are created and their parameters are randomly drawn.

4.2.2 Simulation of games

For each villager, a thread is created to make him play against Minimax *NB_GAMES* times. This saves a lot of time. Once all the games are finished, the parents can be ranked from best to worst. The quality of a parent is defined by the following fraction,

$$\frac{\text{reward}}{\text{nb_simu}}$$

The *reward* is the number of points recovered for all games played and *nb_simu* is the number of games.

4.2.3 Evolution of the population

Once the population has finished playing, merging can begin. *NB_KEEP* parents are kept from the current population (the best) to create the next one. Pairs are then formed and there will be as many pairs as there are individuals missing in order to have the same amount of population at every generation. Each couple formed creates a new child villager through the single point cross-over method.

After merging, one or some children are eventually mutated. The rate of mutations is determined with the global parameter *RATE_MUTATION*. For every new villager, the algorithm picks a random real between 0 and 1, if the real is smaller than the rate, a mutation occurs on the villager. Then a parameter to determine (cf. 4.1) is chosen randomly and a mutation is done. If

it is the number of iterations, it's just increases until it reaches the maximum. If it is one of the factors, the parameter is transform into a binary number and one bit is switched.

4.2.4 Continuation and end

The loop starts again with the newly generated population. Each parent plays a certain number of games, receives a reward, and is allowed to play a game... The algorithm stops when the convergence criterion is checked. ATTENTION ON DOIT CHANGER CE CRITERE D'AILLEURS

5 Benchmark and analysis of the AI

5.1 Initial parameters and approach

Once the algorithm is written, some initial parameters must be choosen. Here : the number of iterations, that we will note n , and the parameter of trade-off between explotation and exploration, that we will note p . As the performance of the AI will depend on those, an initial choice was made, arbitrarily, and depending on the behaviour and performance of the AI, they have been modified.

The most important conclusion that was pointed out was the fact that the iterations must be **as fast as possible**, because the more n can grow, the better. This means that the initial **implementation of the game** must be rethought if it is not optimized for being as fast as possible. Then, during the execution of the algorithm, some **slow methods** must be avoided, e.g. [deepcopy](#). Avoid some redundant calls to functions, etc.

With such cleaning of the algorithm, the AI went from 20 seconds by move with $n = 200$ to 20 seconds by move with $n = 20k$. Note that the documentation about MCTS informs us that having relatively small n , like $n = 200$, will result in an AI playing randomly.

n is a parameter that tells how much the tree is "studied" : we want a tree with the highest n possible for a realistic time. On the other hand, p tells us about the way the tree is going to be used.

- Small p : exploration has a privilege
- Big p : explotation has a privilege

Hence, once n was set to being the biggest possible for realistic execution time, p can be tweaked to have a smarter AI : should we explore more, or exploit more ? This will depend on the AI's performance for each p , and will be discussed in the next sections.

5.2 Benchmark setup and implementation

To analyze the performance of the AI, the following tests were conducted. 4 values of n were tested, and 8 values of p . They can be seen on table 1. For each combination, 10 games were simulated, giving a total of $8 \times 4 \times 10 = 320$ simulations, that took about 26 hours to run. The implementation can be found in the [benchmark.py](#) file.

5.3 Wins and losses

From the 320 simulations, we found nothing to conclude on the performance on the AI on its winning ability. There are 6 wins out of those 320 games, and the rest are draws and losses³

³The initial implementation of Minimax thinks that it is winning as long as it has more pieces left, and the game ends up by drawing.

	p
	0.25
	0.5
n	0.75
5000	1
10000	1.25
15000	1.5
20000	1.75
	2

Table 1: Test values for n and p .

From this, we conclude that no matter the value of p , the MCTS AI can still not find its path to victory, and that n should therefore be improved, by optimizing the run in the initial implementation, or finding another way to speed up the process (as will be discussed later).

5.4 Execution time behaviour depending on the parameters

In this section, we answer to the following questions to understand the reaction of the execution time in result to modifying the parameters :

- How does t react when increasing n ? Linearly ? Exponentially ? This is answered on figure 1.
- Does p have an influence on the execution time ? Answered on figure 2.

A Listings

```

1  class MCNode:
2      def __init__(self, state: Board, color, nb_king_moved, max_it, parent=
None, move: Move = None):
3          self.state: Board = state
4          self.color = color
5          self.adv_color = WHITE if self.color == RED else RED
6          self.reward = 0
7          self.visits = 1 # We always visit newly created node
8          self.parent = parent
9          self.parent_action = move
10         self.children: List[MCNode] = []
11         self.children_moves = []
12         self.max_it = max_it
13         self.nb_king_moved = nb_king_moved # To keep a track to see if the
game needs to end
14         return
15

```

Listing 1: Constructor of a Monte-Carlo Tree-Search Node

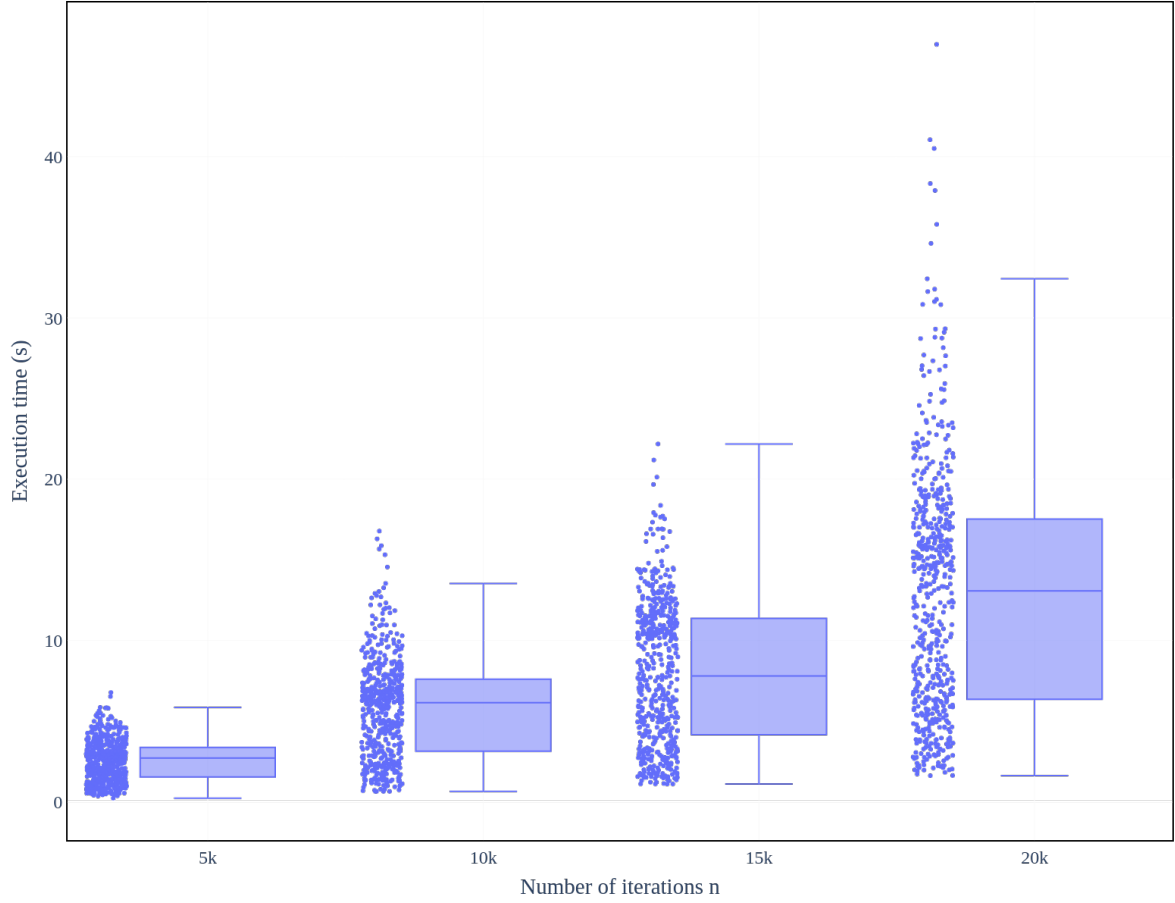


Figure 1: Box-plot of the execution time (in seconds) with several values of n , for $p = 4$. A linear dependance can be seen for the median.

Algorithm 1 Selection of a node. Calls the **Best Child policy** and the **Expansion** procedure.

```

procedure SELECT( $N$ )
     $N$  ▷ MC Root Node
    CurrentNode =  $N$ 
    while CurrentNode is not leaf do
        if CurrentNode is not fully explored then
            return CurrentNode.Expand()
        else
            CurrentNode = BestChild(CurrentNode)
        end if
    end while
end procedure

```

Algorithm 2 Expansion of a node

```

procedure EXPAND( $N$ )
     $N$  ▷ Current MC Node
     $r$  = random(RemainingMoves( $N$ ))
    newBoard =  $N$ .board.simulateMove( $r$ )
     $N_r$  = MCNode( $N$ ,  $r$ , newBoard)
    return  $N_r$ 
end procedure

```

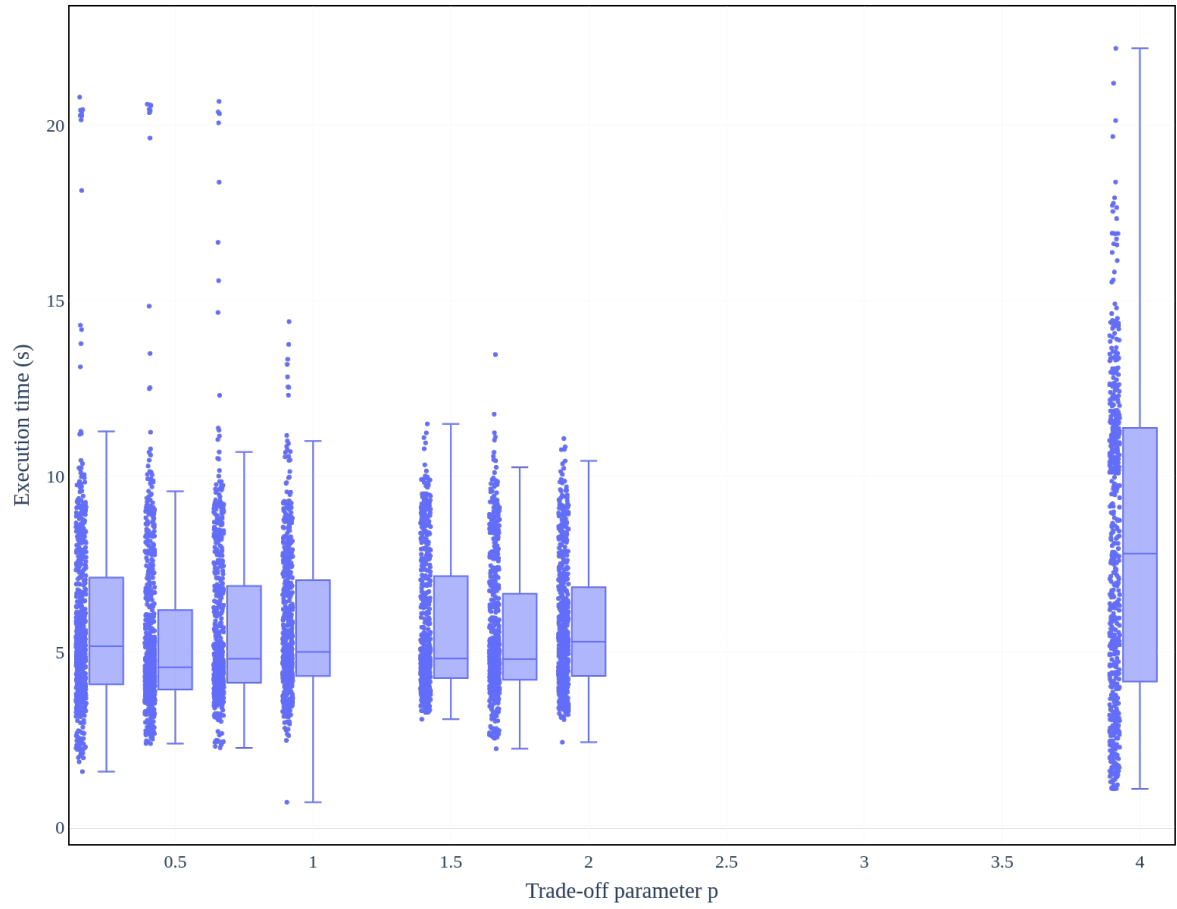


Figure 2: Box-plot of the execution time (in seconds) with several values of p , for $n = 15k$. As expected, the average execution time does not depend on p , hence is constant, although it seems like the execution time seems to spread for high p .