

# Plan van aanpak

Richard Deurwaarder

## Doelstelling / Doel van scriptie

Het doel van de scriptie is helderheid bieden in hoeverre het exploratie vs exploitatie dilemma van toepassing is bij Monte Carlo Tree search en in hoeverre het nodig is om tijd te investeren in het 'tunen' van een oplossing voor dit dilemma bij nieuwe problemen.

## Hoofd- en deelvragen

*Hoofdvraag:*

- Is een optimale c parameter waarde in MCTS UCT significant beter? (de c parameter stuurt exploratie vs exploitatie)

*Deelvragen:*

- Wanneer is een waarde optimaal in een spel van Othello?
- Is er een optimale waarde? Converteert het voor een specifiek probleem?
- Als er een optimale waarde is, geldt deze dan ook voor varianten op het probleem?
- In hoeverre is deze optimale waarde significant? Hoe ver mag een waarde van het 'optimum' af zitten voor het praktische nadelen heeft.

## Onderzoeksmethode / Op welke wijze ga ik onderzoek doen

Stuk software schrijven waarmee ik gemakkelijk experimenten kan uitvoeren. Een los experiment zou moeten bestaan uit 2 spelers die op basis van MCTS-UCT met ieder een aparte waarde voor de c parameter het spel tegen elkaar uitspelen. Hier komt een binaire uitslag uit, win of verlies.

Hier aan toegevoegd is een manier om tot optimale c waarden te komen. Met een klein evolutionair algoritme kan een populatie gemaakt worden van c waarden, die tegen elkaar zullen spelen. Winnaars zullen in de nieuwe populatie komen en een clone van deze zullen gemuteerd worden.

Na een aantal generaties zou het gemiddelde van deze populatie de geconvergeerde waarde moeten zijn.

## Planning

*Al uitgewerkt:*

- Initiële plan van aanpak gemaakt
- Software ontwikkelt
- Initiële literatuur studie om te onderzoeken wat er gezegd wordt over de c parameter en waarom deze in het algoritme / policy zit.
- Ruwe uitlijning van paper opgesteld.

*Nog uit te werken, in chronologische volgorde:*

1. Introductie van paper met achtergrond informatie uitwerken. Uitleg over
  - Othello (99% af)
  - MCTS (90% af)
  - UCT, heeft nog wat extra zoekwerk nodig (10% af)
  - Exploratie vs Exploitatie dilemma
  - Praktische informatie over software / uitvoer experimenten
2. Bepalen hoe ik de experimenten wil gaan uitvoeren:
  - Hoeveel denktijd per zet
  - wat is de range van mogelijke waarden, zijn er waarden die niet kunnen (negatieve waarden?)
  - wanneer is de c waarde geconvergeerd, hoeveel cijfers achter de komma
3. Uitgebreide uitleg over welke experimenten zijn uitgevoerd, waarom deze zijn uitgevoerd (=antwoord op deelvragen) en wat het resultaat is.
  - Initieel wil ik testen of en in welke mate verschillen in c waarden uitmaken. Dit kan getest worden door steeds 2 waarden een paar 100 keer tegen elkaar te laten spelen en kijken wat de win percentages zijn. Op het moment dat dit richting 50% gaat is het verschil niet meer significant.
  - Vervolgens wil ik kijken wat de optimale waarde is voor othello, zonder domain knowledge in de simulatie.
  - Daarna wil ik kijken wat de optimale waarde is voor Othello met domain knowledge en in welke mate deze verschilt van het experiment daarvoor
4. Conclusie op basis van de resultaten / antwoorden op de deelvragen.