

# Othello

— Entwicklung einer KI für das Spiel —

Patrick Müller, Max Zepnik

23. Januar 2019

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>2</b>
<b>2</b>	<b>Grundlagen</b>	<b>3</b>
2.1	Spieltheorie . . . . .	3
2.2	Spielstrategien . . . . .	4
2.2.1	Min-Max . . . . .	4
2.2.2	Alpha-Beta Pruning . . . . .	4
2.2.3	Suboptimale Echtzeitentscheidungen . . . . .	7
2.3	Monte Carlo Tree Search . . . . .	9
2.3.1	Funktionsweise . . . . .	9
2.3.2	Die Selection Policy . . . . .	9
<b>3</b>	<b>Othello</b>	<b>11</b>
3.1	Spielregeln . . . . .	12
3.2	Spielverlauf . . . . .	12
3.3	Spielstrategien . . . . .	13
3.4	Eröffnungszüge . . . . .	13
<b>4</b>	<b>Implementierung der KI</b>	<b>15</b>
<b>5</b>	<b>Evaluierung</b>	<b>16</b>
<b>6</b>	<b>Fazit</b>	<b>17</b>

# Kapitel 1

## Einleitung

Computergegner ..

. test..

text1 ...

am Ende schreiben

auf Fazit beziehen?

# Kapitel 2

## Grundlagen

### 2.1 Spieltheorie

In dem folgenden Unterkapitel werden grundlegende Definitionen eingeführt. Diese sind an [RN16] angelehnt.

**Definition 1 (Spiel (Game))**(vgl. [RN16] S. 162)) Ein **Game** besteht aus einem Tupel der Form

$$\mathcal{G} = \langle S_0, \text{player}, \text{actions}, \text{result}, \text{terminalTest}, \text{utility} \rangle$$

$S_0$  beschreibt den Startzustand des Spiels.

PLAYER ist auf der Menge der Spieler definiert und gibt den aktuellen Spieler zurück.

ACTIONS gibt die validen Folgezustände eines gegebenen Zustands zurück.

RESULT definiert das Resultat einer durchgeführten Aktion  $a$  und in einem Zustand  $s$ .

TERMINALTEST prüft ob ein Zustand  $s$  ein Terminalzustand, also Endzustand, darstellt.

UTILITY gibt einen Zahlenwert aus den Eingabewerten  $s$  ( Terminalzustand) und  $p$  (Spieler) zurück.

Positive Werte stellen einen Gewinn, negative Werte einen Verlust dar.

Definition  
States davor

Eine spezielle Art von Spielen sind **Nullsummenspiele**.

**Definition 2 (Nullsummenspiele)** (vgl. [RN16] S. 161)) In einem **Nullsummenspiel** ist die Summe der utility Funktion eines Zustands über alle Spieler 0. Dies bedeutet, dass wenn ein Spieler gewinnt mindestens ein Gegenspieler verliert.

Durch den Startzustand  $S_0$  und der Funktion ACTION wird ein **Spielbaum (Game Tree)** aufgespannt.

**Definition 3 (Spielbaum (Game Tree))**(vgl. [RN16] S. 162)) Ein **Spielbaum** besteht aus einer **einigen Wurzel**, welche einen bestimmten Zustand (meistens  $S_0$ ) darstellt. Die Kindknoten der Wurzel stellen die durch ACTIONS erzeugten Zustände dar. Die Kanten zwischen der Wurzel und den Kindknoten stellen jeweils die durchgeführte Aktion dar, die ausgeführt wurde um vom State  $s$  zum Kindknoten zu gelangen.

**Definition 4 (Suchbaum (Search Tree))**(vgl. [RN16] S. 163)) Ein **Suchbaum** ist ein Teil des Spielbaums.

Überleitung  
einfügen

## 2.2 Spielstrategien

Es gibt verschiedene Spielstrategien. Im Folgenden werden diese kurz erläutert und anschließend verglichen.

### 2.2.1 Min-Max

Der erste hier erläuterte Strategie ist der Min-Max Algorithmus. Dieser ist folgendermaßen definiert:

Zitat einfügen

$$MinMax(s) = \begin{cases} Utility(s); & \text{wenn TerminalTest}(s) == \text{true} \\ \max(\{a_e \text{ Actions}(s) MinMax(Result(s, a))\}); & \text{wenn Spieler am Zug} \\ \min(\{a_e \text{ Actions}(s) MinMax(Result(s, a))\}); & \text{wenn Gegner am Zug} \end{cases}$$

Der Spieler sucht den bestmöglichen Zug aus ACTIONS, der ihm einen für seine Züge einen Vorteil schafft aber gleichzeitig nur „schlechte“ Zugmöglichkeiten für den Gegner generiert. Der Gegner kann dadurch aus allen ehemals möglichen Zügen nicht den optimalen Zug spielen, da dieser in den aktuell enthaltenen Zügen nicht vorhanden ist. Er wählt aus den verfügbaren ACTIONS nach den gleichen Vorgaben seinen besten Zug aus.

Die Strategie ist eine Tiefensuche und erkundet jeden Knoten zuerst bis zu den einzelnen Blättern bevor ein Nachbarknoten ausgewählt wird. Dies setzt das mindestens einmalige Durchlaufen des gesamten Search Trees voraus. Bei einem durchschnittlichen Verzweigungsfaktor von  $f$  bei einer Tiefe von  $d$  resultiert daraus eine Komplexität von  $O(d^f)$ . Bei einem einmaligen Erkunden der Knoten können die Werte aus den Blättern rekursiv von den Blättern zu den Knoten aktualisiert werden. Dadurch muss im nächsten Zug nur das Minimum aus ACTIONS ermittelt werden, da alle Kindknoten schon evaluiert wurden. Für übliche Spiele kann die Min-Max-Strategie allerdings nicht verwendet werden, da die Komplexität zu hoch für eine akzeptable Antwortzeit ist und der benötigte Speicherplatz für die berechneten Zustände sehr schnell wächst.

### 2.2.2 Alpha-Beta Pruning

Der Min-Max Algorithmus berechnet nach dem Prinzip „depth-first“ stets den kompletten Game Tree. Bei der Betrachtung des Entscheidungsverhaltens des Algorithmus fällt jedoch schnell auf, dass ein nicht unerheblicher Teil aller möglichen Züge gar nicht erst in Betracht gezogen wird. Dies geschieht aufgrund der Tatsache, dass diese Züge in einem schlechteren Ergebnis resultieren würden als die letztendlich ausgewählten.

Dem Alpha-Beta Pruning Algorithmus liegt der Gedanke zugrunde, dass die Zustände, die in einem realen Spiel nie auftreten würden auch nicht berechnet werden müssen. Damit steht die dafür regulär erforderliche Rechenzeit und der entsprechende Speicher dafür zur Verfügung andere, vielversprechendere Zweige zu verfolgen.

#### Demonstration an einem Beispiel

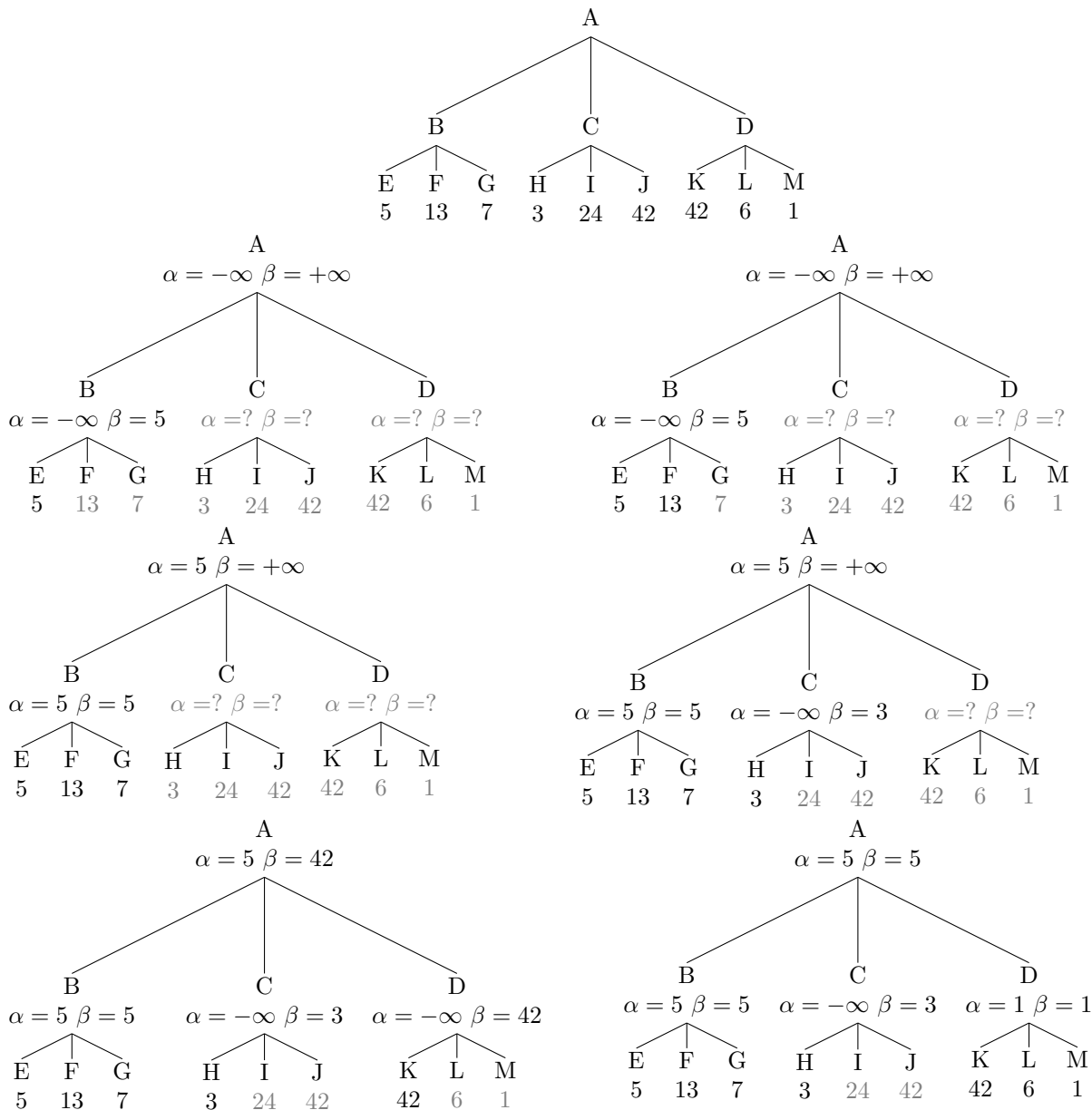
Um den Algorithmus zu verdeutlichen betrachten wir das, an [RN16] angelehnte, folgende Beispiel. Das dargestellte Spiel besteht aus lediglich zwei Zügen, die abwechselnd durch die Spieler gewählt werden. An den Knoten der untersten Ebene des Game Tree werden die Werte der Zustände gemäß der UTILITY Funktion angegeben. Die Werte  $\alpha$  und  $\beta$  geben den schlecht möglichsten bzw. den bestmöglichen Spielausgang für einen Zweig, immer aus der Sicht des beginnenden Spielers, an. Die ausgegrauten Knoten wurden noch nicht betrachtet.

Betrachten wir nun den linken Baum in der zweiten Zeile: Der Algorithmus beginnt damit alle möglichen Folgezustände bei der Wahl von B als Folgezustand zu evaluieren. Dabei wird zunächst der Knoten E betrachtet und damit der Wert 5 ermittelt. Dies ist der bisher beste Wert. Er wird als  $\beta$  gespeichert. Eine Aussage über den schlechtesten Wert kann noch nicht getroffen werden.

Im nachfolgenden Game Tree wird der nächste Schritt verdeutlicht. Es wird der Knoten F betrachtet. Dieser hat

Warum

Abbildung 2.1: Beispielhafter Game Tree



einen Wert von 13. Am Zuge ist jedoch der zweite Spieler. Dieser wird, geht man davon aus, dass er ideal spielt, jedoch keinen Zug wählen der ein besseres Ergebnis für den Gegner bringt als unbedingt nötig. Der bestmögliche Wert für den ersten Spieler bleibt damit 5.

Nach der Auswertung des Knotens G steht fest, dass es keinen besseren und keinen schlechteren Wert aus Sicht des ersten Spielers gibt. Daraufhin wird die 5 auch als schlechtester Wert in  $\alpha$  gespeichert. Ausgehend von A ist der schlechteste Wert damit 5 ggf. kann jedoch noch ein besseres Ergebnis herbeigeführt werden.  $\alpha$  wird entsprechend gesetzt und  $\beta$  verbleibt undefiniert.

Nun werden die Kindknoten von C betrachtet. Mit einem Wert von 3 wäre der Knoten H das bisher beste Ergebnis für die Wahl von C. Der Wert wird entsprechend gespeichert. Würde C gewählt gäbe man dem Ge-

genspieler die Chance ein im Vergleich zu der Wahl des Knotens B schlechteres Ergebnis herbeizuführen. Da Ziel des Spielers jedoch ist, die eigenen Punkte zu maximieren, gilt es diese Chance gar nicht erst zu gewähren. Entsprechend werden die Auswertung der weiteren Knoten abgebrochen.

Der Kindknoten K des Knotens D ist mit einem Wert von 42 vielversprechend und wird in  $\beta$  gespeichert. Da dieser Wert größer ist als die gespeicherten 5 wird auch der entsprechende Wert von A aktualisiert. Der anschließend ausgewertete Knoten L ermöglicht nun ein schlechteres Ergebnis von 6  $\beta$ , muss also aktualisiert werden. Der Knoten M liefert schließlich den schlechtesten Wert von 1. Da der Gegenspieler im Zweifel diesen Wert wählen würde, bleibt der bisher beste Wert das Ergebnis in E. In A wird der Spieler daher B auswählen.

Dieses einfache Beispiel zeigt bereits recht gut, wie die Auswertung von weiteren Zweigen vermieden werden kann. In der Praktischen Anwendung befinden sich die wegfallenden Zustände häufig nicht nur in den Blättern des Baumes, sondern auch auf höheren Ebenen. Der eingesparte Aufwand wird dadurch häufig noch größer.

## Implementierung

Nachfolgend wird eine Pseudoimplementierung des um Alpha-Beta Pruning erweiterten MinMax Algorithmus angegeben (siehe Listing 2.1):

Listing 2.1: Pseudoimplementierung von Alpha-Beta Pruning

```

1 global Suchtiefe
2 int minMax(Spiel AktuellerZustand, int Spieler, int Tiefe, int alpha, int beta) {
3     if (Tiefe == 0) {
4         return Utility(AktuellerZustand, Spieler);
5     }
6     int bisherigerMaximalWert = alpha
7     Zuege = mengeDerFolgezuege(aktuellerZustand);
8     for (Zug z in Zuege) {
9         Spiel NeuerZustand = waehleZug(Aktueller_Zustand, z)
10        wert = -minMax(NeuerZustand, anderer(Spieler), Tiefe-1, -beta, -bisherigerMaximalWert)
11        if (wert > bisherigerMaximalWert) {
12            bisherigerMaximalWert = wert
13            if (bisherigerMaximalWert >= beta) {
14                break;
15            }
16            if (Tiefe = Suchtiefe) {
17                speichereZug(z)
18            }
19        }
20    }
21    return bisherigerMaximalwert;
22 }
```

Es handelt sich um eine rekursive Implementierung. Im Basisfall ist der Game Tree bereits bis in die angegebene Suchtiefe erforscht (Zeile 3). In diesem Fall wird der Wert der Utility Funktion für den aktuellen Spieler bei dem aktuellen Zustand zurückgegeben (Zeile 4).

Handelt es sich nicht um einen solchen Fall, werden alle möglichen Folgezüge berechnet (Zeile 7) und dann **einzelnen betrachtet in dem er ausgeführt wird**. (Zeile 8f). Zuvor wird dazu jedoch der bisherige Maximalwert gespeichert (Zeile 6). Um den Wert des Zuges zu bestimmen wird rekursiv die minMax-Methode erneut aufgerufen. Dabei wird entsprechend der Neue Zustand, der andere Spieler und eine um die um eins verringerte Tiefe

übergeben. Der beste Wert für den anderen Spieler ist der schlechteste Wert für den ersten Spieler. Daher wird der bisherige Wert von beta als alpha übergeben. Der bisher beste Wert ist aus Sicht des anderen Spielers der schlechteste, daher wird dieser als neues beta übergeben. Da die Utility Funktion so implementiert ist, dass die Summe der Wertigkeiten eines Zustandes Null ergibt, muss noch das Vorzeichen geändert werden (Zeile 9). Ist der neue Wert größer als der bisherige Maximalwert (Zeile 11), so wird dieser aktualisiert (Zeile 12). Da der zweite Spieler versucht die Punktzahl des Gegners zu maximieren, bricht dieser die Auswertung aller Zweige ab, bei denen ein Ergebnis, welches besser ist als das bisher schlechteste Ergebnis, möglich wird (Zeile 13f). Abschließend wird der ausgewertete Zug gespeichert um ihn später ausführen zu können (Zeile 16).

### Ordnung der Züge

Wie in obigen Beispiel an den Zweigen unter dem Knoten C zu sehen war kann, je nach der Reihenfolge in der die Folgezüge untersucht werden, die Auswertung eines Folgezustandes früher oder später abgebrochen werden. Optimalerweise werden die besten Züge, also jene Züge die einen möglichst frühen Abbruch der Betrachtung eines Knotens herbeiführen zuerst betrachtet. Um dies Abschätzen zu können bedient man sich in der Praxis einer Heuristik die Aussagen über die Güte eines Zuges im Vergleich zu den übrigen Zügen zulässt. Anhand dieser Heuristik kann dann die Reihenfolge der Auswertung einzelner Folgezustände dynamisch angepasst werden.

### 2.2.3 Suboptimale Echtzeitentscheidungen

Selbst die gezeigte Verbesserung des MinMax-Algorithmus besitzt noch einen wesentlichen Nachteil. Da es sich um einen "depth-first" Algorithmus handelt muss jeder Pfad bis zu einem Endzustand betrachtet werden um eine Aussage über den Wert des Zuges treffen zu können. Dem steht jedoch die Tatsache entgegen, dass in der Praxis eine Entscheidung möglichst schnell, idealer Weise innerhalb weniger Minuten, getroffen werden soll. Hinzu kommt, dass je nach der verwendeten Datenstruktur für ein Spiel bei entsprechend hohem Verzweigungsfaktor und einer großen Anzahl von Zügen der Hauptspeicher eines handelsüblichen Computers nicht mehr ausreicht um diese zu fassen

Es gilt also eine Möglichkeit zu finden, die Auswertung des kompletten Baumes zu vermeiden.

### Heuristiken

Dieses Problem lösen sogenannte Heuristiken. Dabei handelt es sich um eine Funktion die den Wert eines Spielzustandes annähert.

Die Nutzung der Heuristik wird vereinfacht, wenn Sie so definiert ist, dass sie, sofern es sich um einen Endzustand handelt den Wert der Utility Funktion zurückgibt. Der Vorteil dieses Verhaltens wird im nächsten Abschnitt betrachtet.

Kommt eine Heuristik zur Anwendung, so ist die Genauigkeit, mit der diese den tatsächlichen Wert approximiert, der wesentliche Aspekt, der die Qualität des Spiel-Algorithmus ausmacht. Um zu verhindern, dass versehentlich die besten Züge nicht betrachtet werden, ist es essentiell, dass eine Heuristik den tatsächlichen Wert eines Zustandes nie überschätzt. Das unterschätzen des Wertes hingegen ist möglich darf im Sinne der Genauigkeit der Heuristik jedoch nicht allzu ungleichmäßig auftreten.

admissible?

consistent?

### Abschnittskriterium der Suche

Gibt die Heuristik im Falle eines Endzustandes den Wert der Utility Funktion zurück, so kann die oben gezeigte Implementierung so angepasst werden, dass statt der Utility Funktion einfach die Heuristik ausgewertet wird.



Dadurch muss nicht mehr der Vollständige Zweig durchsucht werden und das Abbrechen nach einer gewissen Suchtiefe wird möglich.

### Forward pruning

Forward pruning durchsucht nicht den kompletten **Game Tree**, sondern durchsucht nur einen Teil. Eine Möglichkeit ist eine Strahlensuche, welche nur die „besten“ Züge durchsucht (vgl. [RN16] S. 175). Die Züge mit einer geringen Erfolgswahrscheinlichkeit werden abgeschnitten und nicht bis zum Blattknoten evaluiert. Durch die Wahl des jeweils wahrscheinlichsten Zuges können aber auch sehr gute bzw. schlechte Züge nicht berücksichtigt werden, **da sie eine geringe Wahrscheinlichkeit besitzen**. Durch das Abschneiden von Teilen des Game Tree wird die Suchgeschwindigkeit deutlich erhöht. Der in dem Othello-Programm „Logistello“ verwendete „Probcut“ erzielt außerdem eine Gewinnwahrscheinlichkeit von 64% gegenüber der ursprünglichen Version ohne Forward pruning (vgl. [RN16] S. 175).

### Search versus lookup

Viele Spiele kann man in 3 Haupt-Spielabschnitte einteilen:

- Eröffnungsphase
- Mittelspiel
- Endphase

In der Eröffnungsphase und in der Endphase gibt es im Vergleich zum Mittelspiel wenige Zugmöglichkeiten. Dadurch sinkt der Verzweigungsfaktor und die generelle Anzahl der **states**. In diesen Phasen können die optimalen Spielzüge einfacher berechnet werden. Eine weitere Möglichkeit besteht aus dem Nachschlagen des Spielzustands aus einer Lookup-Tabelle.

Dies ist sinnvoll, da gewöhnlicherweise sehr viel Literatur über die Spieleröffnung des jeweiligen Spiels existiert. Das Mittelspiel jedoch hat zu viele Zugmöglichkeiten, um eine Tabelle der möglichen Spielzüge bis zum Spielende aufstellen zu können. In dem Kapitel 3.4 werden die bekanntesten Eröffnungsstrategien aufgelistet.

Viele Spielstrategien wie beispielsweise die Min-Max-Strategie setzen den kompletten oder wenigstens einen großen Teil des Spielbaums voraus. Dieser kann entweder berechnet werden oder aus einer Lookup-Tabelle gelesen werden. Je nach Verzweigungsfaktor der einzelnen Spielzüge kann diese allerdings sehr groß sein. Selbst im späten Spielverlauf gibt es verschiedene Spiele, welche einen großen Spielbaum besitzen.

Beispielsweise existieren für das Endspiel in Schach mit einem König, Läufer und Springer gegen einen König 3.494.568 mögliche Positionen (vgl. [RN16] S.176).

Dies sind zu viele Möglichkeiten um alle speichern zu können, da noch sehr viel mehr Endspiel-Kombinationen als diese existieren.

Anstatt die Spielzustände also zu speichern können auch die verbleibenden Spielzustände berechnet werden. Othello besitzt gegenüber Schach den Vorteil, dass die Anzahl der Spielzüge auf 60 bzw. 64 Züge begrenzt sind. Dadurch kann in der Endphase des Spiel ggf. der komplette verbleibende Game Tree berechnet werden, da die Anzahl der möglichen Zugmöglichkeiten eingeschränkt wird.

Bei der Berechnung der Spielzüge sind die Suchtiefe und der Verzweigungsfaktor entscheidend für die Berechnungsdauer. Aus diesem Grund können im Mittelspiel keine Min-Max-Algorithmen bis zu den Blattknoten des Game Trees ausgeführt werden, da die Menge des benötigten Speicherplatzes außerhalb jeglicher Grenzen eines Arbeits- oder Gamingcomputers liegen.

## 2.3 Monte Carlo Tree Search

Die sogenannte Monte-Carlo Tree Search (MCTS - Monte-Carlo Baumsuche) bedarf im Gegensatz zu den bisher gezeigten Strategien in ihrer Reinform keine Heuristik. Dabei werden zufällige Spiele gespielt. Aus einem einzigen solchen Spiel lässt sich kaum eine Erkenntnis ableiten; aus einer Vielzahl von zufälligen Spielen lässt sich jedoch bei einer ausreichend großen Anzahl die optimale Lösung bestimmen.

### 2.3.1 Funktionsweise

[CBSS08] beschreiben das Verfahren als einen vierstufigen Prozess zum Aufbau eines Game Trees.

#### Selection

Ist der Ausgangszustand bereits bekannt, also bereits im Game Tree enthalten, so wird der Folgezustand aufgrund der vorhandenen Daten gewählt. In der Regel stehen hier solche Folgezustände zu denen bereits Daten vorhanden sind und solche, die bisher unbekannt sind zur Verfügung. Die Schwierigkeit liegt nun darin zu entscheiden, ob jener Folgezustand der am vielversprechendsten ist (exploitation) oder ein bisher unbekannter Zustand der unter Umständen ein besseres Ergebnis liefern könnte (exploration) gewählt wird. Die bei der Auswahl angewandte Strategie wird als Selection Policy bezeichnet. Genauer wird diese Problemstellung im Abschnitt 2.3.2 behandelt.

#### Expansion

Wenn ein Zustand erreicht wird, der bisher nicht im Game Tree enthalten ist, so wird dieser hinzugefügt. Durch die folgenden beiden Schritte werden dann Informationen zu diesem Zustand gespeichert.

#### Simulation

Nun werden bis zum Erreichen eines Terminalzustandes zufällige Züge durchgeführt. In weiteren Optimierungen kann hier eine Heuristik eingeführt werden um vielversprechende Züge zuerst zu erkunden.

#### Backpropagation

Im letzten Schritt werden dann die gespeicherten Informationen durch Backpropagation angepasst. Dabei wird die Häufigkeit des Besuchs eines Zustandes, sowie jeweils die Häufigkeit eines Gewinn bzw. Verlusts bei der Wahl dieses Zustandes gespeichert. Der Wert des Zustandes kann nun durch die Anzahl der Gewinne bei Wahl der Aktion durch die Besuchshäufigkeit angenähert werden.

Nach dem derartigen Aufbau des Game Trees wurde aufgrund der Auswahlbedingung im Selection-Schritt jener Folgezustand am häufigsten erkundet, der am Erfolgsversprechendsten ist. Im tatsächlichen Spiel wird daher der Zug durchgeführt, der beim Aufbau des Baumes am häufigsten durchgeführt wurde.

### 2.3.2 Die Selection Policy

Das Problem des Auswählens des durchzuführenden Zuges ist analog zu dem sogenannten K-Armed Bandit Problem. Dabei spielt der Spieler an einem mehrarmigen Banditen, also einem Glücksspielautomaten. Bei der Wahl eines Armes wird mit einer bestimmten Wahrscheinlichkeit ein Gewinn ausgeschüttet. Damit steht der Spieler vor jedem Zug vor der Wahl: Er kann entweder den Arm betätigen oder nach seinem Wissen den höchsten

Gewinn verspricht, geht dabei aber das Risiko ein, einen Arm der einen bedeutend höheren Gewinn ermöglicht nicht zu betätigen, oder er kann einen Arm zu dem ihm bisher noch keine Informationen vorliegen spielen um ggf. einen Arm mit besseren Chancen zu finden. In der Literatur wird dieses Problem als Exploration-Exploitation Dilemma bezeichnet.

Das Problem kann als eine Reihe von unabhängigen Zufallsvariablen  $X_{i,n}$  betrachtet werden. Dabei steht  $1 \leq i \leq K$  für den Arm des Banditen und  $n \geq 1$  für den Zug. Das Spielen eines Armes  $i$  ergeben die Gewinne  $X_{i,1}, X_{i,2}, \dots, X_{i,n}$  die gemäß einer zunächst unbekannten Vorschrift mit dem Erwartungswert  $\mu_i$  berechnet wird. [BPW<sup>+</sup>12]

Nachfolgend finden sich einige Strategien um die Wahl des Armes bzw. Folgezustandes vorzunehmen:

### $\epsilon$ -Greedy

Die  $\epsilon$ -greedy Policy ist eine vergleichsweise einfache Variante zur Lösung des Problems. Um der exploration Rechnung zu tragen wird dabei mit einer festen Wahrscheinlichkeit  $\epsilon$  ein zufälliger Zug ausgewählt. Andernfalls kommt jener Zug zum Einsatz der den nach aktuellem Wissensstand höchsten Gewinn verspricht. In einer angepassten Variante, vorgeschlagen durch [Tok10], wird die Wahrscheinlichkeit  $\epsilon$  je nach Wissensstand des Spielers angepasst. Zu **beginn** ist sie damit bspw. vglw. hoch während sie bei zunehmender Sicherheit verringert wird.

### Regret

Die Regret-Policy versucht den Verlust durch die Wahl eines anderen als den nach derzeitigem Stand optimalen Schrittes so gering wie möglich zu halten. Dieser Verlust wird für  $n$  Durchgänge wie folgt berechnet:

$$R_N = \mu^* n - \mu_j \sum_{j=1}^K E[T_j(n)]$$

Dabei steht  $\mu^*$  für den erwarteten Maximalgewinn und  $E[T_j(n)]$  für die erwartete Anzahl der Züge bei denen der Arm  $j$  gewählt wurde.

### Upper Confidence Bound

[ACBF02] haben gezeigt, dass es eine Strategie, von ihnen Upper Confidence Bound 1 (UCB1) genannt, gibt, die ein logarithmisches Wachstum des Regretts über  $n$  ermöglicht, ohne dass dazu weitere Informationen bezüglich der Gewinnverteilung bekannt sein müssen, sobald die Belohnungen zwischen 0 und 1 liegen. Die Strategie spielt dabei jenen Arm  $j$ , der UCB1 maximiert, mit:

$$UCB1 = \bar{X}_j + \sqrt{\frac{2 \ln n}{n_j}}$$

Dabei ist  $\bar{X}_j$  der durchschnittliche Gewinn beim Spielen des Armes  $j$ ,  $n_j$  die Anzahl der Male zu denen  $j$  gewählt wurde und  $n$  die Anzahl der insgesamt gespielten Durchgänge. Der linke Term ist der durchschnittliche Gewinn und stellt damit die Exploitation sicher, während sich der rechte Term für selten gewählte Arme stetig erhöht und die Exploration sicherstellt.

# Kapitel 3

## Othello

Othello wird auf einem 8x8 Spielbrett mit zwei Spielern gespielt. Es gibt je 64 Spielsteine, welche auf einer Seite schwarz, auf der anderen weiß sind. Der Startzustand besteht aus einem leeren Spielbrett, in welchem sich in der Mitte ein 2x2 Quadrat aus abwechselnd weißen und schwarzen Steinen befindet. Anschließend beginnt der Spieler mit den schwarzen Steinen.

Die Spielfelder werden in verschiedene Kategorien eingeteilt (siehe Abbildung 3.1):

- Randfelder: äußere Felder (blaue Felder) [o.V15]
- C-Felder: Felder, welche ein Feld horizontal oder vertikal von den Ecken entfernt sind (vgl. [Ber])
- X-Felder: Felder, welche ein Feld diagonal von den Ecken entfernt sind [o.V15]
- Zentrum: innerste Felder von C3 bis F6 (grüne Felder) [o.V15]
- Zentralfelder: Felder D4 bis E5 [o.V15]
- Frontsteine: die äußersten Steine auf dem Spielbrett um das Zentrum (vgl. [Ort]).

Diese Kategorien sind für die spätere Strategie wichtig.

	A	B	C	D	E	F	G	H
1		C					C	
2	C	X					X	C
3								
4				W	S			
5				S	W			
6								
7	C	X					X	C
8		C					C	

Abbildung 3.1: Kategorien des Spielfeldes

Von Othello gibt verschiedene Varianten. Eine Variante ist Reversi. Die verschiedenen Varianten sind allerdings bis auf die Startposition gleich. Bei der Variante Reversi sind die Zentralfelder noch nicht besetzt und die Spieler setzen die vier Steine selbst, während bei Othello die Startaufstellung fest vorgegeben ist.

### 3.1 Spielregeln

Othello besitzt einfache Spielregeln, welche im Spielverlauf aber auch taktisches oder strategisches Geschick erfordern. Jeder Spieler legt abwechselnd einen Stein auf das Spielbrett. Dabei sind folgende Spielregeln zu beachten welche auch in Abbildung 3.2 abgebildet sind:

- Ein Stein darf nur in ein leeres Feld gelegt werden.
- Es dürfen nur Steine auf Felder gelegt werden, welche einen oder mehrere gegnerischen Steine mit einem bestehenden Stein umschließen würden. Dies ist im linken Spielbrett durch grüne Felder und im mittleren Spielbrett durch das gelbe Feld hervorgehoben. Das gelbe Feld (F5) umschließt mit dem Feld D5 (blau) einen gegnerischen Stein. Es können auch mehrere Steine umschlossen werden. Allerdings dürfen sich dazwischen keine leeren Felder befinden.
- Von dem neu gesetzten Stein in alle Richtungen ausgehend werden die umschlossenen gegnerischen Steine umgedreht, sodass alle Steine die eigene Farbe besitzen. In dem Beispiel ist das im dem rechten Spielbrett zu sehen. F5 umschließt dabei das Feld E5 (rot). Dieses Feld wird nun gedreht und wird schwarz.
- Ist für einen Spieler kein Zug möglich muss dieser aussetzen. Ein Spieler darf allerdings nicht freiwillig aussetzen wenn noch mindestens eine Zugmöglichkeit besteht.
- Ist für beide Spieler kein Zug mehr möglich, ist das Spiel beendet. Der Spieler mit den meisten Steinen seiner Farbe gewinnt das Spiel.
- Das Spiel endet auch wenn alle Felder des Spielbrettes besetzt sind. In diesem Fall gewinnt ebenfalls der Spieler mit den meisten Steinen seiner Farbe.

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6								
7								
8								

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6								
7								
8								

	A	B	C	D	E	F	G	H
1								
2								
3								
4								
5								
6								
7								
8								

Abbildung 3.2: valide Zugmöglichkeiten für Schwarz und ausgeführter Zug

### 3.2 Spielverlauf

Das Spiel wird in drei Abschnitte eingeteilt [Ort]:

- Eröffnungsphase
- Mittelspiel
- Endspiel

Diese Abschnitte sind jeweils 20 Spielzüge lang. Im Eröffnungs- und Endspiel stehen zum Mittelspiel wenige Zugmöglichkeiten zur Verfügung, da entweder nur wenige Steine auf dem Spielbrett existieren oder das Spielbrett fast gefüllt ist und nur noch einzelne Lücken übrig sind. Im Mittelspiel existieren sehr viele Möglichkeiten, da sich schon mindestens 20 Steine auf dem Spielbrett befinden und diese sehr gute Anlegemöglichkeiten bieten.

### 3.3 Spielstrategien

Wie in anderen Spielen gibt es auch in Othello verschiedene Strategien. Dabei kann beispielsweise offensiv gespielt werden, indem versucht wird möglichst viele Steine in einem Zug zu drehen. Es gibt auch defensive „stille“ Züge. Ein „stiller“ Zug dreht keinen Frontstein um und dreht möglichst nur wenige innere Steine um (vgl. [Ort]).

Generell ist eine häufig genutzte Strategie die eigene Mobilität zu erhöhen und die Mobilität des Gegners zu verringern. Mit dem Begriff Mobilität sind die möglichen Zugmöglichkeiten gemeint. Durch das Einschränken der gegnerischen Mobilität hat dieser weniger Zugmöglichkeiten und muss so ggf. strategisch schlechtere Züge durchführen.

Die Position der Steine auf dem Spielbrett sollte ebenfalls nicht vernachlässigt werden. beispielsweise sollen Züge auf X-Felder vermieden werden, da der Gegner dadurch Zugang zu den Ecken bekommt. Dadurch können ggf. die beiden Ränder und die Diagonale gedreht werden und in den Besitz des Gegners gelangen.

In der Eröffnungsphase sollten die Randfelder ebenfalls vermieden werden, da diese in dieser frühen Phase des Spiels noch gedreht werden können und der taktische Vorteil in einen strategischen Nachteil umgewandelt wird.

### 3.4 Eröffnungszüge

In der nachfolgenden Tabelle 3.1 sind verschiedene Spieleröffnungen und deren Häufigkeit in Spielen aufgelistet. Spielzüge werden in Othello durch eine Angabe der Position, auf welche der Stein gesetzt wird, dargestellt. Ein vollständiges Spiel lässt sich deshalb in einer Reihe von maximal 60 Positionen darstellen.

Name	Häufigkeit	Spielzüge
Tiger	47%	F5 D6 C3 D3 C4
Rose	13%	F5 D6 C5 F4 E3 C6 D3 F6 E6 D7
Buffalo	8%	F5 F6 E6 F4 C3
Heath	6%	F5 F6 E6 F4 G5
Inoue	5%	F5 D6 C5 F4 E3 C6 E6
Shaman	3%	F5 D6 C5 F4 E3 C6 F3

Tabelle 3.1: Liste von Othelloeröffnungen [Ort]

[Ort] gibt folgende weitere Tipps für Eröffnungen:

- Versuche weniger Steinchen zu haben als dein Gegner.
- Versuche das Zentrum zu besetzen.

- Vermeide zu viele Frontsteine umzudrehen.
- Versuche eigene Steine in einem Haufen zu sammeln statt diese zu verstreuen.
- Vermeide vor dem Mittelspiel auf die Kantfelder zu setzen.

Viele dieser Tipps können auch im späteren Spielverlauf verwendet werden.

## Kapitel 4

# Implementierung der KI











## Kapitel 5

# Evaluierung

## Kapitel 6

## Fazit

# Notes

	am Ende schreiben . . . . .	2
	auf Fazit beziehen? . . . . .	2
	Definition States davor . . . . .	3
	Überleitung einfügen . . . . .	3
	Zitat einfügen . . . . .	4
	Warum . . . . .	4
	admissible? . . . . .	7
	consistent? . . . . .	7

# Literaturverzeichnis

- [ACBF02] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
- [Ber] Berg, Matthias. Strategieführer. <http://berg.earthlingz.de/ocd/strategy2.php>. [Online; accessed 20-January-2019].
- [BPW<sup>+</sup>12] Cameron B. Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1):1–43, 2012.
- [CBSS08] Guillaume Chaslot, Sander Bakkes, Istvan Szita, and Pieter Spronck. Monte-carlo tree search: A new framework for game ai. In *AIIDE*, 2008.
- [Ort] Ortiz, George and Berg, Matthias. Eröffnungsstrategie. <http://berg.earthlingz.de/ocd/strategy3.php>. [Online; accessed 20-January-2019].
- [o.V15] o.V. Spiele: Othello. [https://de.wikibooks.org/wiki/Spiele:\\_Othello](https://de.wikibooks.org/wiki/Spiele:_Othello), 2015. [Online; accessed 20-January-2019].
- [RN16] Stuart J. Russell and Peter Norvig. *Artificial intelligence: A modern approach*. ~~Always learning.~~ Pearson, ~~Boston and Columbus and Indianapolis and New York and San Francisco and Upper Saddle River and Amsterdam, Cape Town and Dubai and London and Madrid and Milan and Munich and Paris and Montreal and Toronto and Delhi and Mexico City and Sao Paulo and Sydney and Hong Kong and Seoul and Singapore and Taipei and Tokyo,~~ third edition, ~~global edition edition~~, 2016.
- [Tok10] Michel Tokic. Adaptive ~~\epsilon~~-greedy exploration in reinforcement learning based on value differences. In Dillmann, Rüdiger and Beyerer, Jürgen and Hanebeck, Uwe D. and Schultz, Tanja, editor, *KI 2010: Advances in Artificial Intelligence*, pages 203–210, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.