# **Projektbericht:**

# **Maschinelles Lernen am Spiel „Vier-Gewinnt“**

### Business Analytics Project

### Jonathan Cawalla, Lena Gräwe, Ahmad Haschemi und Lena Knickmeier

### Betreut durch:

### Dr. Michael Baumann, Dr. Thomas Spura

Inhaltsverzeichnis

1. **Einleitung3**
2. **Umsetzung der Trainingsumgebung3**
3. **Q-Learning6**
4. **Neuronal Netze8**
5. **Test und Auswertung10**
6. **Fazit und Ausblick21**
7. **Verweise23**

## 1. Einleitung

Im Rahmen des Business Analytics Projekt sollte anlehnend an die aktuellen Erfolge von Google mit ihrer künstlichen Intelligenz (KI) *AlphaGo* eine lernende KI implementiert werden. Die *AlphaGo-*KI weist im komplexen Spiel „Go“ eine Gewinnquote von 99,8% vor und schlug den amtierenden Europameister. [vgl. 1] Für das Business Analytics Projekt bestand die Aufgabe darin, ein vergleichsweise leichtes Spiel auszuwählen und eine bzw. zwei Methoden des Maschinellen Lernens darauf anzuwenden, um eine KI zu implementieren.

Als erster Schritt sollte ein klassisches Spiel ausgewählt werden, welches ein niedrigeres Komplexitätsniveau als Go hat. Die Auswahl fiel dabei auf das Spiel „Vier-Gewinnt“. Vier-Gewinnt ist ein Spiel bei dem zwei Spieler abwechselnd einen Stein auf dem Spielfeld platzieren. Das Spielfeld besteht aus sieben Spalten, in denen je sechs Steine Platz haben. Spielsteine können nur von oben in die Spalten geworfen werden und fallen dann soweit nach unten wie möglich. Ein Spieler hat gewonnen, wenn er es schafft, vier seiner Steine direkt nebeneinander zu platzieren. Dabei ist es egal, ob sie horizontal, vertikal oder auf einer Diagonalen liegen.

Es gibt mehrere Möglichkeiten, um die Komplexität eines Spiels zu messen. Eine Möglichkeit ist, den Spielbaum eines Spiels zu betrachten. Jedes Spiel lässt sich durch einen zusammenhängenden und schleifenlosen Graph bestehend aus Knoten und Ästen abbilden. Für Entscheidungsknoten (mit nach unten weiterführenden Ästen) ist anzugeben, welcher Spieler hier seinen Zug gewählt hat. Der Wert eines Blattes bestimmt, ob der Spieler gewonnen oder verloren hat. Zur Evaluation der möglichen Züge wird ein Spielbaum rekursiv durchgesucht, der aus schätzungsweise möglichen Sequenzen von Zügen besteht, wobei b die Breite des Spiels (Anzahl der legalen Züge pro Position oder Verzweigungsfaktor) darstellt und d ist Tiefe des Spiels (Spieldauer). In den großen Spielen wie Schach gilt b≈35, d ≈80 und für Go gilt b ≈250, d ≈150 [vgl. 1]. Für das Spiel Vier-Gewinnt gilt b≈4, d≈36 [vgl. 2].

Ein anderer Maßstab für die Komplexität eines Spiels ist die Anzahl der möglichen Spielsituationen im Spiel. Für das klassische Vier-Gewinnt auf einem Spielfeld mit 6 Reihen und 7 Spalten gibt es 4.531.985.219.092 mögliche Positionen für alle Situationen von 0 bis 42 Steine im Spiel [vgl. 3].

## 2. Umsetzung der Trainingsumgebung

**Spielumgebung**

Um Zeit und Arbeit zu sparen, sollte mit einer fertigen Implementierung aus dem Internet begonnen werden. Das entsprechende Programm sollte bestenfalls leicht anpassbar sein und bereits eine einfache KI enthalten.

Nach ausgiebiger Recherche und kleinen Experimenten mit einigen der gefunden Programmen fiel die Auswahl auf eine Implementierung von Vier-Gewinnt die zunächst brauchbar schien [vgl. 4]. Um das Programm einsetzbar zu machen, wurden einige Änderungen vorgenommen, um die Spielfeldgröße über globale Variablen einstellen zu können. Außerdem sollte ein dritter Spielmodus eingeführt werden, um zwei KIs gegen einander spielen lassen zu können. Die Umsetzung dieses Modus barg jedoch mehr Tücken als ursprünglich angenommen. Das größte Problem bestand darin, dass, die Reihenfolge der Spieler vertauscht wurde. D.h. immer, wenn Spieler 1 an der Reihe sein sollte, wurde ein Stein von Spieler 1 positioniert und umgekehrt.

Da der Fehler nicht behoben werden konnte und nicht klar war in wie weit er sich auf den Lernerfolg der KIs auswirken würde, fiel letzten Endes der Entschluss, die bisherige Arbeit zu verwerfen und eine eigene Version von Vier-Gewinnt zu programmieren.

Die Klasse *Game* enthält jeweils eine globale Variable für die Spielfeldbreite, eine für die Spielfeldhöhe und eine für die Anzahl von Steinen, die nebeneinander gereiht werden müssen, um zu gewinnen. Das Spielfeld selbst wird in einem zwei dimensionalen Array *board* gespeichert, auf das mit einem Getter zugegriffen werden kann. Um die Implementierung von Spiel und KIs zu erleichtern, enthält die Spielklasse außerdem je eine Indikator-Variable, um anzuzeigen, ob das Spielbrett leer ist und ob das Spiel beendet wurde.

Neben einigen Hilfsmethoden mit denen u.a. geprüft wird, ob das Spielende erreicht ist, das Spielbrett zurücksetzt werden kann oder Steine auf dem Spielfeld platziert werden, enthält das Spiel vier unterschiedliche Methoden für die Durchführung von Spielen:

* *PlayGame*Hierbei handelt es sich um eine Methode, die mit zwei übergebenen Spielern ein einzelnes Spiel durchführt und das Ergebnis auf der Konsole ausgibt.
* *generateDataset*Die Methode bekommt zwei Spieler und eine Anzahl an durchzuführenden Spielen übergeben und gibt nach entsprechend vielen Spielen ein Datenset zurück.
* *trainQPlayer*Die Methode bekommt zwei Spieler und die Anzahl zu spielender Spiele übergeben. Außerdem erhält sie eine Information darüber, ob die Spieler abwechselnd beginnen sollen oder nicht.
* *playTournament*Die Methode erhält die gleichen Eingabe-Parameter wie *trainQPlayer*. Auch die Funktionsweise ist im Wesentlichen gleich, mit dem Unterschied, dass falls ein Q-Player am Turnier teilnimmt, keine Werte in seiner Datenbank mehr verändert werden. Diese Methode dient somit vor allem zur Beurteilung der Trainingsergebnisse. Auf der Konsole werden Spielzüge, Ergebnisse von einzelnen Spielen sowie summierte Kennziffern über das gesamte Turnier ausgegeben.

**NormalKI und NormalKI2**

Damit die KIs später automatisiert lernen können, wird eine weitere KI benötigt, die bereits spielfähig ist. Das oben erwähnte Vier-Gewinnt-Spiel enthält bereits eine gute KI, deren Algorithmus für die neue Spiel-Implementierung übernommen wurde.

Der enthaltene Algorithmus basiert auf einfachen if-then-Abfragen und Schleifen. Dabei wird zunächst geprüft, ob das Spielfeld leer ist. In diesem Fall wirft sie in die mittlere Spalte, um eine möglichst gute Ausgangslage zu schaffen. Wenn bereits Steine auf dem Spielfeld liegen, prüft die KI als erstes, ob sie selbst gewinnen kann. Falls nicht, wird abgefragt, ob der Gegner gewinnen kann und der eigene Stein entsprechend so geworfen, um dies zu verhindern.

Sofern keiner dieser Fälle vorliegt, wird eine defensive Spielweise gewählt. Für ein klassisches Vier-Gewinnt-Spiel wird nun geprüft, welche Züge es dem Gegner erlauben würden drei bzw. zwei seiner Steine nebeneinander zu platzieren. Alle entsprechenden Züge werden als Spalten in einem Array *possibleSolutions* zwischengespeichert. Da Züge, die es ermöglichen drei Steine neben einander zu platzieren, natürlich auch zwei Steine neben einander zur Folge haben, werden sie doppelt in das Array aufgenommen. Anschießend werden alle Züge aus *possibleSolutions* entfernt, die es dem Gegner in seinem nächsten Zug ermöglichen (über eine Diagonale oder Horizontale) zu gewinnen. Im nächsten Schritt werden alle verbleibenden Spalten, die nicht am Rand liegen, ein weiteres Mal in das Array kopiert, um dann einen zufälligen Zug auszuwählen. Falls *possibleSolutions* leer ist, wird nach einer Lösung gesucht, die in erster Linie erlaubt ist und dem Gegner nach Möglichkeit nicht ermöglicht, im nächsten Zug zu gewinnen.

Die Klasse *NoramlKI2* enthält außerdem eine Variable *epsilon*, die festlegt wie viel Prozent der Spielzüge rein zufällig ausgewählt werden sollen. Dies wurde äquivalent zur im nächsten Abschnitt folgenden epsilon-greedy-Strategie beim *QPlayer2* implementier [vgl. 9] und soll dafür sorgen, dass die KIs eine größere Auswahl an Spielzügen kennen lernen. Dies ist der einzige Unterschied zu *NormalKI.*

**Datenbank**

Da die Anzahl der möglichen Spielzustände sehr hoch ist, wird eine effiziente Speicherstruktur für den Q-Player sehr wichtig. Es müssen beliebig viele Zustände gespeichert werden können und der Zugriff auf die Zustände muss sehr schnell ablaufen, da der Q-Player während eines Spieldurchlaufs in jedem Zug sowohl für diesen Zug, als auch für die nächsten Züge Abfragen an die Datenbank generiert. Aus diesem Grund wurde eine HashMap gewählt, die intern von der Klasse *Q\_DB* verwendet wird. Der Schlüssel dieser HashMap ist ein Spielzustand. Der Spielzustand wird von der Klasse *Game* als 2-dimensionales Int Array gespeichert. Da dieser Datentyp sich aufgrund einer fehlenden Hash- und Equalsfunktion nicht als Schlüssel eignet, wird jeder Spielzustand in einen *Array2DWrapper* umgewandelt, der diese Funktionen implementiert. Als Wert für die HashMap wird wiederrum eine neue Datenstruktur verwendet, die eine Aktion und den Wert für diese Aktion enthält. Die Aktion ist die Spalte in die geworfen wird und der Wert ist abhängig von der internen Bewertung des Q-Players und wird von diesem verändert. Die dafür verwendete Datenstruktur ist ebenfalls eine HashMap. Es könnte auch eine ArrayList oder ähnliches verwendet werden, da hier nur eine maximale Anzahl von sieben Aktionen und damit sieben Einträgen möglich ist.

Die Datenbank implementiert weiterhin die folgenden Methoden mit denen der Q-Player seine Werte abfragen und verändern kann:

* *get(..)*, *containsState(..)* für Abfragen an die Datenbank
* *getValueOfStateAndAction(..)*: Liefert für einen bestimmten Zustand und die Aktion, die darauf ausgeführt werden soll, den Wert dieser Aktion zurück.
* *put(..)* und *update(..)* zum Verändern der Datenbank.
* Sowie Funktionen zum Speichern und Laden der Datenbank über die Serialisierung der HashMap.

Das Speichern und Laden der HashMap wurde insbesondere dafür verwendet, die Datenbank für den Q-Player nach einer gewissen Anzahl an Trainingsiterationen zu speichern. Damit kann das Training zum späteren Zeitpunkt fortgesetzt werden, indem die Datenbank wieder geladen wird.

## Q-Learning

**Methode**

Beim Q-Learning handelt es sich um eine einfache Form des Reinforcement Learning. Das Prinzip basiert darauf, dass der KI in jedem Zustand eine Reihe von Aktionen zur Auswahl stehen. Je nachdem, welche Aktion sie wählt, erhält sie eine Belohnung oder Bestrafung, die sich zum einen Teil aus dem unmittelbar reichten Zustand und zum andern aus den später erreichbaren Zuständen ableiten lassen. Indem die KI versucht ihre Belohnungen zu maximieren, lernt sie über mehrere Durchläufe, was die bestmögliche Spielstrategie ist. [vgl. 5]

Im Falle des Spiels Vier-Gewinnt definieren sich die Zustände durch die aktuelle Position aller Steine auf dem Spielbrett. Die Aktion der KI ist das Werfen eines Spielsteins in eine der Spalten.

**Verwandte Arbeiten**

Olszowka beschäftig sich in seiner Arbeit neben einer allgemeinen Implementierung von Q-Learning für Vier-Gewinnt mit den damit verbundenen Schwierigkeiten, die ihm in einer früheren Arbeit begegnet sind. Unter anderem geht er auf die große Anzahl von Spielzuständen beim klassischen Vier-Gewinnt ein und beschäftigt sich mit Methoden, um die Anzahl der für den Algorithmus relevanten Zustände zu reduzieren. [vgl. 6]

Șimșek und Barto stellen neben viele anderen mit Q-Learning gelösten Problemen eine KI für Tic-Tac-Toe vor. [vgl. 7] Tic-Tac-Toe ist zwar auf Grund der kleineren Spielfeldgröße weniger komplex als Vier-Gewinnt, die Spielweise ist jedoch ähnlich.

**Implementation des Q-Players**

Für die Umsetzung des Q-Players diente eine Anleitung aus dem Internet als grobe Orientierung. [vgl. 8]

Für die Verteilung der Werte der einzelnen Zustände diente nach Absprache mit dem Betreuer die folgende Formel als Ausgangspunkt:

Dabei ist *alpha* der Prozentsatz zudem der alte Wert aus der Datenbank bestehen bleiben soll. Es wurde ein Wert von 0,95 gewählt, sodass eine 5 % Veränderung möglich ist, sollte der Q-Player, denselben Zustand noch einmal betreten und den Wert verändern wollen. Das verhindert, dass der Wert jedes Mal überschrieben wird, sobald er neu berechnet wird. Stattdessen konvergiert dieser mit steigender Lernzeit gegen einen festen Wert. Gamma ist der Lernparameter mit dem gewichtet wird, wie stark zukünftige Züge in die Bewertung einfließen. Es wurde ein Wert von 0,8 gewählt. Die Belohnung wird verteilt, sobald das Spiel gewonnen ist.

Um dieses Verfahren an das Vier-Gewinnt Spiel anzupassen, wurden mehrere Veränderungen gemacht. Es wurde nicht nur eine Belohnung für ein gewonnenes Spiel, sondern auch eine Bestrafung für ein verlorenes Spiel gesetzt. Diese werden immer dann verteilt, wenn ein Spiel abgeschlossen ist. Der Q-Player wird über die Methode *reactToWinOrLose* darüber informiert und schreibt den neuen Wert in die Datenbank. Dies passiert getrennt von der Berechnung der restlichen Formel, da immer dann, wenn ein Spiel abgeschlossen ist, die Berechnung von nicht mehr nötig ist, da es keine zukünftigen Züge gibt.

Die Formel wurde für das 2-Spieler-Spiel angepasst. Im ursprünglichen Q-Learning wird ein 1-Spieler-Spiel betrachtet, sodass der Spieler nach einem Zug direkt wieder am Zug ist und daher direkt die nächsten Züge aus der Datenbank gelesen werden können. Bei Vier-Gewinnt muss erst der gegnerische Spieler seinen Stein platzieren, bevor der Q-Player selbst wieder an der Reihe ist. Die Berechnung erfolgt in der Methode *turn* und wurde wie folgt angepasst:

wird in der Methode *avgValueForNextStateAllActions* über eine Simulation der nächsten möglichen Züge berechnet. Zunächst werden alle möglichen Züge für den Q-Player generiert. Das Generieren von Zügen wird ermöglicht durch die Methode *generateActions*. Basierend auf diesen Zügen werden die nächsten Züge für den Gegner berechnet. Es wird nur eine Heuristik verwendet, um einen dieser Züge auszuwählen. Dafür wird der Spielzustand invertiert und in der Datenbank des Q-Players geschaut, ob dieser Züge für diesen Spielzustand vorhanden sind. Falls ja, dann wird derjenige mit dem höchsten Wert ausgewählt. Damit wird nur die bestmögliche Aktion für den Gegner betrachtet, was sinnvoll erscheint, da dieser immer diese Aktion machen würde und nicht eine beliebige andere. Falls nein, wird ein zufälliger Zustand ausgewählt. Anschließend werden für diese ausgewählte Aktion des Gegners alle möglichen Züge des Q-Players generiert. Es wird der Durchschnitt der Werte all dieser Aktionen gebildet und in gespeichert. Der Durchschnitt wird aus dem Grund gebildet, da ansonsten die Bestrafungen nicht mit in den Q-Wert einbezogen werden.



Abbildung

Um zu verhindern, dass sich der Q-Player auf eine unausgereifte Strategie festfährt, wurde zusätzlich ein epsilon-greedy-Stategie angewandt. Dabei erhält der Q-Player einen weiteren Parameter ɛ, der festlegt, zu wie viel Prozent der Q-Player seine Aktionen im Training zufällig wählt. [vgl. 9]

## Neuronales Netz

**Methode**

In Sinne des Maschinellen Lernens handelt es sich bei künstlichen Neuronalen Netzen um eine Struktur, welche aus Neuronen und die Neuronen miteinander verbindenden Synapsen besteht, ähnlich dem Zentralnervensystem von Lebewesen. Künstlich Neuronale Netze werden für Einschätzungen und Approximationen von Funktionen angewendet, die mit einer großen Menge von Input-Elementen abhängig sind. [vgl. 10] Diese Neuronalen Netze können durch drei Eigenschaften festgelegt werden: der Architektur, der Aktivierungsregel und den Lernregeln.

Die Architektur bestimmt, welche Variablen im Neuronalen Netz und seiner Topologie beteiligt sind. Die Anzahl der Neuronen in den verdeckten Schichten (Hidden-Layer) sind beispielweise variierbar, im Gegensatz zu der Eingabe- und Ausgabeschicht (Input- und Output-Layer). Es wurde in dieser Arbeit ein Multi-Layer Perzeptron mit einer verdeckten Schicht (Hidden-Layer) ausgesucht.

Die Aktivierungsregel bestimmt, wie sich die Aktivitäten der Neuronen in Reaktion aufeinander verändern. Z. B. verfügt jede verdeckte Schicht und die Ausgabeschicht über eine (eigene) Aktivierungsfunktion. Diese können linear oder nicht linear sein. Nicht lineare Aktivierungsfunktionen machen das Neuronale Netz besonders mächtig. [vgl. 11]

In diesem Projekt wurde die Sigmoidfunktion als Aktivierungsfunktionen angewendet, da der Einsatz von differenzierbaren Funktionen die Verwendung von Lernmechanismen (Backpropagation-Algorithmus) ermöglicht. Als Aktivierungsfunktion eines Neurons wird die Sigmoidfunktion auf die Summe der gewichteten Eingabewerte angewendet, um die Ausgabe des Neurons zu erhalten. Die Sigmoidfunktion wird vor allem aufgrund ihrer einfachen Differenzierbarkeit als Aktivierungsfunktion bevorzugt verwendet. [vgl. 11]

Die Lernregeln bestimmen die Art und Weise, in der die Gewichte des Netzes durch Training im Laufe der Zeit angepasst werden. Es wurde hier Backpropagationals Lernalgorithmus angewendet, welches ein verbreitetes Verfahren für das Erlernen von künstlichen neuronalen Netzen ist. [vgl. 12]

**Verwandte Arbeiten**

Schneider und Rosa haben sich damit beschäftigt vier heuristische Algorithmen zu implementieren, die relativ fähig Vier-Gewinnt spielen können. Diese nutzten sie, um Daten-Sets für das Training der Neuronalen Netze zu generieren. Außerdem haben sie sich damit auseinandergesetzt, die Qualität eines Netzes durch Variation der Parameter sowie Netzwerk Topologie und der Größe der Trainingsmenge zu verbessern. [vgl. 13]

Chellapilla und Fogel haben sich damit befasst, den Ansatz der künstlichen Intelligenz durch Implementierung von KIs für ein Nullsummenspiel (Tic-Tac-Toe) und ein nicht Nullsummenspiel (Gefangenendilemma) auszuprobieren. Dafür haben sie zwei Methoden des Maschinellen Lernens, nämlich Neuronale Netze und Evolutionärer Algorithmus, angewendet, und miteinander verglichen. [vgl. 14]

**Implementation des NN-Players**

Nach ersten Recherchen zum Thema Neuronale Netze fiel die Entscheidung auf das Framework „Neuroph 2.7“ von SourceForge, welches eine Java Bibliothek als auch die GUI-Anwendung „Neuroph Studio“ bietet. Mit der Java Bibliothek können Neuronale Netze in Java Programmen erzeugt, trainiert und angewendet werden [vgl. 15]. Es gibt außerdem eine Anleitung zur Erzeugung von Multi-Layer Perzeptronen [vgl. 16], die für das weitere Vorgehen hilfreich war, weil sich an dem Vorgehen von Schneider und Rosa [vgl. 13] orientiert wurde. Für das Neuronale Netz wurde die Struktur von Multi-Layer Perzeptron mit einem Hidden-Layer angewandt. Für das Lernen wurde Backpropagation und die Sigmoid-Funktion als Aktivierungsfunktion gewählt, wie auch bei Schneider und Rosa [vgl. 13]. Danach berechnet sich die Anzahl der Neuronen im Input-Layer als Produkt aus Spalten- und Zeilen-Anzahl, multipliziert mit drei, denn jede Zelle des Spielfelds wird mit drei Eigenschaften beschreiben. Die Anzahl ist also abhängig von der Spielfeldgröße. Eine genauere Erläuterung folgt im nächsten Abschnitt. Die Anzahl der Neuronen im Hidden-Layer kann durch Ausprobieren ermittelt werden. Im späteren Training und Testverlauf wurden mehrere Neuronale Netze mit unterschiedlicher Neuronenanzahl im Hidden-Layer erstellt und deren Auswirkung erfasst. Dabei wurde die Anzahl der Input-Neuronen mit einem Faktor i multipliziert, wobei für Faktor i unterschiedliche Werte gewählt wurden (s. Kapitel 5). Die Anzahl der Neuronen im Output-Layer ergibt sich aus der Anzahl der Spalten des jeweiligen Spielfelds.

Damit das Neuronale Netz lernen kann, werden Daten benötigt. Hierfür wurde die Methode *generateDataSets* in der Klasse *Game* implementiert. Durch Aufruf dieser in der *main*-Methode der Klasse *Game* wird eine bestimmte Anzahl von Spielen zweier *IPlayer* ausgeführt und deren jeweiligen Zustände in eine Textdatei geschrieben, wobei hier zu beachten ist, dass nur neue, unbekannte Zustände in die Datei geschrieben werden. Außerdem werden nur die Zustände von einem bestimmten Spieler und nicht vom gegnerischen Spieler gespeichert.

Bei der Generierung der Textdatei als Daten-Set werden die jeweiligen Spielzustände in Input- und Output-Elemente übersetzt, die das Neuronale Netz später einlesen und im Lernprozess verarbeiten kann. Hierbei wurde die Idee von Schneider und Rosa übernommen, die Input-Elemente binär darzustellen. [vgl. 13] So wird ein einzelnes Feld auf dem Spielfeld durch eine Folge von drei Ziffern ausgedrückt, deren Wert beim Zutreffen eines bestimmten Merkmals eins und bei Nicht-Zutreffen null ist.

Die erste Stelle dieser Ziffern-Folge beschreibt, ob das Feld leer ist, die zweite Stelle, ob dieses Feld von einem Stein des Spieler 1 belegt und die dritte Stelle, ob das Feld von einem Stein des Spieler 2 belegt wird. So steht die Folge [1, 0, 0] für ein leeres Feld, die Folge [0, 1, 0] für ein Feld mit einem Stein des Spieler 1 und die Folge [0, 0, 1] für ein Feld mit einem Stein des Spieler 2. Eine Zeile in der Textdatei einspricht einem Spielzustand (Input für das Neuronale Netz) und die Aktion (Output für das Neuronale Netz) darauf, d. h. es gibt bspw. für ein 6x7 Feld eine Ziffernfolge von 133 Ziffern, wobei die ersten 126 Ziffern den Spielzustand (Input) und die letzten 7 Ziffern die entsprechende Aktion (Output) darstellen. Die sieben Ziffern der Aktion bestehen aus einer Folge von Nullen und Einsen. Diese Folge beschreibt mit einer Eins, in welche Spalte der Stein geworfen wurde, und füllt die restlichen sechs Spalten mit Nullen auf, damit eine eindeutige Spalte zugewiesen werden kann. Ein solches Dataset kann später im Lernprozess des Neuronalen Netzes eingelesen und mit diesen Input-Elementen trainiert werden.

Damit das Neuronale Netz für eine Spielteilnahme genutzt werden kann, wurde die Klasse *NNPlayer2* implementiert. Im Konstruktor dieser Klasse muss manuell angegeben werden, ob durch Aufruf der Methode *learnNNPlayer* ein neues Neuronales Netz erzeugt und trainiert wird oder ob ein schon vorhandenes Netz geladen wird, indem der Name dieses Netzes eingegeben wird.

In der *learnNNPlayer*-Methode wird ein zuvor erstelltes Daten-Set zum Importieren eingegeben und ein auf die jeweilige Spielsituation angepasstes Multi-Layer Perzeptron wird erzeugt. Hierbei können auch passende Werte für den maximalen Fehler, die Lernrate und das Momentum angegeben werden. Das Neuronale Netz wird nach Generierung abgespeichert, damit es zu einem späteren Zeitpunkt geladen werden kann.

Im Spielverlauf selbst wird die *turn*-Methode der Klasse *NNPlayer2* aufgerufen, um einen Spielzug durchzuführen. Hier wird der aktuelle Spielzustand als Input an das an den *NNPlayer2* geknüpfte Netz übergeben. Das Netz kalkuliert daraus dann einen Output und gibt zurück in welche Spalte der nächste Stein geworfen werden soll.

## Test und Auswertung

Um die Funktion der KIs einfacher prüfen zu können, wurden Training und Tests für beide KIs zunächst auf einem 4x5 Feld als „Drei-Gewinnt“ durchgeführt, bevor der Prozess für ein klassisches 6x7 „Vier-Gewinnt“-Feld wiederholt wurde. Für beide Spieler diente die NormalKI als Gegner. Im Folgenden wird zunächst auf die Methode des Q-Learning eingegangen und danach auf die Methode der Neuronalen Netze.

**Q-Learning 4x5**

Der *QPlayer2* wurde mit einem *epsilon* von 20% und einem *alpha* von 0,05 schrittweise in insgesamt 128.000 Spielen trainiert. Damit der *QPlayer2* möglichst viele Spielzüge kenne lernt warf auch die Normal KI mit einer Wahrscheinlichkeit von 20% rein zufällig. Der anschließende Test wurden in fünf Turnieren mit je 10.000 Spielen durchgeführt.

Abbildung 2

Wie zu erwarten nimmt die Anzahl der Datenbank Elemente mit jedem Training zu, die Anzahl der unbekannten Spielfeldzustände je 10.000 Spiele sinkt. Beides geschieht mit abnehmender Änderungsrate. [siehe Abb. 2]

Abbildung 3

Die Lernkurve steigt zunächst stark an und erreicht ihren Höhepunkt bei 32.000 Trainingsspielen. In diesem Stadium gewinnt sie bei abwechselndem Spielbeginn bis zu 81,8% (auf einer Basis von 10.000 Spielen) und 100% (auf einer Basis von 100.000), wenn sie bei allen Spielen den ersten Zug machen darf. Überraschenderweise führte jeder Versuch, sie weiter zu trainieren, zu einer Verschlechterung der Spielperformance. Dabei wurden sowohl die Parameter als auch die Trainingsintervalle geändert. [siehe Abb. 3]

Hinzuzufügen ist, dass die Anzahl der unentschiedenen Spiele im Schnitt bei 2,56 von 10.000 Spielen liegt und somit vernachlässigbar ist. Genauso liegt der Unterschied zwischen maximal und minimal gewonnen Spielen pro Trainingsstufe im Schnitt bei nur 0,71%.

Besonders interessant ist, dass die KI eine Art strategisches Vorausdenken zeigt, indem sie Zwickmühlen baut. Sie bringt den Gegner also in eine Lage, in der er seine Niederlage nicht mehr verhindern kann. Die folgenden drei Spielausschnitte stammen vom Q-Player mit 32.000 Trainingsspielen und sollen dieses Verhalten veranschaulichen. (*QPlayer2* = ■ *NormalKI*=X)

|0|0|0|0|0| |0|0|0|0|0| |0|0|0|0|0| |0|0|0|0|0|

|0|0|X|0|0| |0|0|X|0|0| |0|0|X|0|0| |■|0|X|0|0|

|0|0|■|0|0| |0|■|■|0|0| |X|■|■|0|0| |X|■|■|0|0|

|■|X|■|X|0| |■|X|■|X|0| |■|X|■|X|0| |■|X|■|X|0|

*Abbildung 4*

Der *QPlayer2* hat hier eine Situation geschaffen, in der er in der zweiten Zeile sowohl links als auch rechts seinen dritten Stein platzieren kann, um zu gewinnen. Dazu kommt, dass der Gegner, wenn er seinen Stein in die linke Spalte wirft, dem *QPlayer2* die Gewinnmöglichkeit über die Diagonale gibt. [siehe Abb. 4]

|0|0|0|0|0| |0|0|0|0|0| |0|0|0|0|0| |0|0|0|0|0|

|0|X|X|0|0| |0|X|X|0|0| |0|X|X|0|0| |0|X|X|0|0|

|0|■|■|0|0| |0|■|■|0|0| |0|■|■|X|0| |0|■|■|X|0|

|0|X|■|0|0| |0|X|■|■|0| |0|X|■|■|0| |0|X|■|■|■|

Abbildung 5

In diesem Spiel positioniert der *QPlayer2* seine Steine so, dass er sowohl in der untersten als auch in der darüber liegenden Zeile eine Dreierreihe vervollständigen kann. [siehe Abb. 5]

|0|X|0|0|0| |0|X|0|0|0| |0|X|0|0|0| |0|X|0|■|0|

|0|X|■|0|0| |0|X|■|0|0| |0|X|■|X|0| |0|X|■|X|0|

|0|■|X|0|0| |0|■|X|■|0| |0|■|X|■|0| |0|■|X|■|0|

|X|■|X|■|0| |X|■|X|■|0| |X|■|X|■|0| |X|■|X|■|0|

*Abbildung 6*

Auch hier hat der *QPlayer2* im zweiten gezeigten Zustand zwei Möglichkeiten zu gewinnen. Zum einen in der vierten Spalte, zum anderen über eine Diagonale von Spalte drei bis fünf. Zusätzlich ermöglicht der Gegner eine weitere Option, wenn er in die vierte Spalte wirft. [siehe Abb. 6]

**Q-Learning 6x7**

Den *QPlayer2* für ein 6x7 Spielfeld zu trainieren ist hingegen sehr viel schwerer. Zwar lernt der Spieler erkennbar. Jedoch benötigt das Training bei dieser Spielfeldgröße deutlich mehr Spiele; mit 3,2 Mio. Trainingsspielen liegt die Performance der KI hier gerade einmal bei durchschnittlichen 11,2%. [siehe Abb. 7]

Abbildung 7

Das eigentliche Problem stellt hier jedoch nicht die Anzahl der benötigten Trainingsspiele dar, sondern die große Anzahl an Spielzuständen und Datenbank-Elementen. [siehe Abb. 8] Nach 1,6 Mio. Spielen enthält die Datenbank gut 11 Mio. Elemente und ist 4,3 GB groß. Nach 3,2 Mio. Spielen belief sich die Anzahl der Elemente auf ca. 21 Mio. Eine genaue Speichergröße liegt nicht vor, da die Datenbank auf Grund fehlender Arbeitsspeicherkapazität nicht gespeichert werden konnte. Eine Schätzung auf 8,2 GB scheint aber realistisch.

Ein weiteres Training auf 4,8 Mio. bzw. 6,4 Mio. Trainingsspielen wurde zwar versucht, jedoch abgebrochen, da der Trainingsprozess zunehmend langsamer wurde. Zuletzt lag die Anzahl der durchgeführten Spiele bei ca. 200 pro Minute, womit das Training mehre Tage gedauert hätte. Es ist sogar denkbar, dass es mehrere Wochen gedauert hätte, da die Geschwindigkeit mit Sicherheit weiter abgenommen hätte.

Abbildung 8

Trotz der geringen Erfolgsquote nach 3.2 Mio. Lerndurchläufen lassen sich einige geschickte Züge in den Turnierspielen finden. Genauso wie auf dem 4x5 Feld konstruiert der *QPlayer2* Zwickmühlen, um seinen Gegner zu besiegen.

|  |  |  |  |
| --- | --- | --- | --- |
| |0|0|0|0|0|0|0|  |0|0|0|0|0|0|0|  |0|0|0|X|0|0|0|  |0|0|0|■|0|0|0|  |0|■|0|■|0|0|0|  |X|X|■|X|X|0|0| | |0|0|0|0|0|0|0|  |0|0|0|0|0|0|0|  |0|0|0|X|0|0|0|  |0|0|0|■|0|0|0|  |0|■|■|■|0|0|0|  |X|X|■|X|X|0|0| | |0|0|0|0|0|0|0|  |0|0|0|0|0|0|0|  |0|0|0|X|0|0|0|  |0|0|0|■|0|0|0|  |X|■|■|■|0|0|0|  |X|X|■|X|X|0|0| | |0|0|0|0|0|0|0|  |0|0|0|0|0|0|0|  |0|0|0|X|0|0|0|  |0|0|0|■|0|0|0|  |X|■|■|■|■|0|0|  |X|X|■|X|X|0|0| |

*Abbildung 9*

Abbildung 9 zeigt eine einfache Form der Zwickmühle, bei der der *QPlayer2* eine Reihe aus drei seiner Steine mit zwei offenen Enden erzeugt. Ähnliche Spielsituationen finden sich mehrfach in den aufgezeichneten Turnierausschnitten.

|  |  |  |  |
| --- | --- | --- | --- |
| |0|0|X|0|0|0|0|  |0|0|■|0|0|0|0|  |0|X|■|X|0|0|0|  |0|■|■|■|0|0|0|  |0|■|X|X|0|0|0|  |X|X|■|X|0|0|0| | |0|0|X|0|0|0|0|  |0|0|■|■|0|0|0|  |0|X|■|X|0|0|0|  |0|■|■|■|0|0|0|  |0|■|X|X|0|0|0|  |X|X|■|X|0|0|0| | |0|0|X|0|0|0|0|  |0|0|■|■|0|0|0|  |0|X|■|X|0|0|0|  |0|■|■|■|0|0|0|  |X|■|X|X|0|0|0|  |X|X|■|X|0|0|0| | |0|0|X|0|0|0|0|  |0|0|■|■|0|0|0|  |0|X|■|X|0|0|0|  |■|■|■|■|0|0|0|  |X|■|X|X|0|0|0|  |X|X|■|X|0|0|0| |

*Abbildung 10*

Im zweiten Beispiel ist eine komplexere Strategie zu sehen. Der *QPlayer2* positioniert zunächst drei seiner Steine auf einer Diagonalen. Der Gegner versucht im nächsten Zug den Sieg des Q-Players zu verhindern und wirft seinen Stein in die linke Spalte. Dadurch wird zwar der Sieg über die Diagonale für den *QPlayer2* verhindert, gleichzeitig ist es ihm nun aber möglich eine Vierer-Reihe in der Horizontalen zu vervollständigen.

**NNPlayer2 4x5**

Zunächst wurden für ein Spielfeld der Größe 4x5 im „Drei-Gewinnt“-Spiel mehrere Datasets erstellt, sowohl für Spiele des 1. Spielers (Beginner eines Spiels) als auch Spiele des 2. Spielers. Die Zeileneinträge der Daten-Sets wurden jeweils auf 200, 300 und 400 Einträge beschränkt, wobei nur unterschiedliche Einträge gespeichert werden, die aber zufällig in einem Turnier von zwei *NormalKI*-Spielern, die gegeneinander spielten, erzeugt wurden. Es wurden so insgesamt sechs Datasets erzeugt, jeweils drei pro Spieler. Es wurden mehrere Neuronale Netze erzeugt, die auf den zuvor generierten Datasets trainiert wurden. Dabei lag maximale Fehler bei 0,01, die bei Lernrate 0,2 und das Momentum bei 0,7. Um die Auswirkung der Neuronen-Anzahl im Hidden-Layer näher zu betrachten, wurden dafür unterschiedlich Anzahlen übergeben, indem die Anzahl der Neuronen im Input-Layer mit einem Faktor i= {½, 1, 2, 3, 4} multipliziert wurden. Dies entspricht einer Neuronen-Anzahl von 30, 60, 120, 180 und 240. Bei dieser Vorgehensweise wurde sich an der Arbeit von Schneider und Rosa orientiert [vgl. 13].

Im weiteren Verlauf wurden mehrere Turniere des *NNPlayer2* mit den unterschiedlichen Neuronalen Netzen gegen die *NormalKI* gespielt, wobei pro Turnier 10.000 Spiele gespielt wurden.

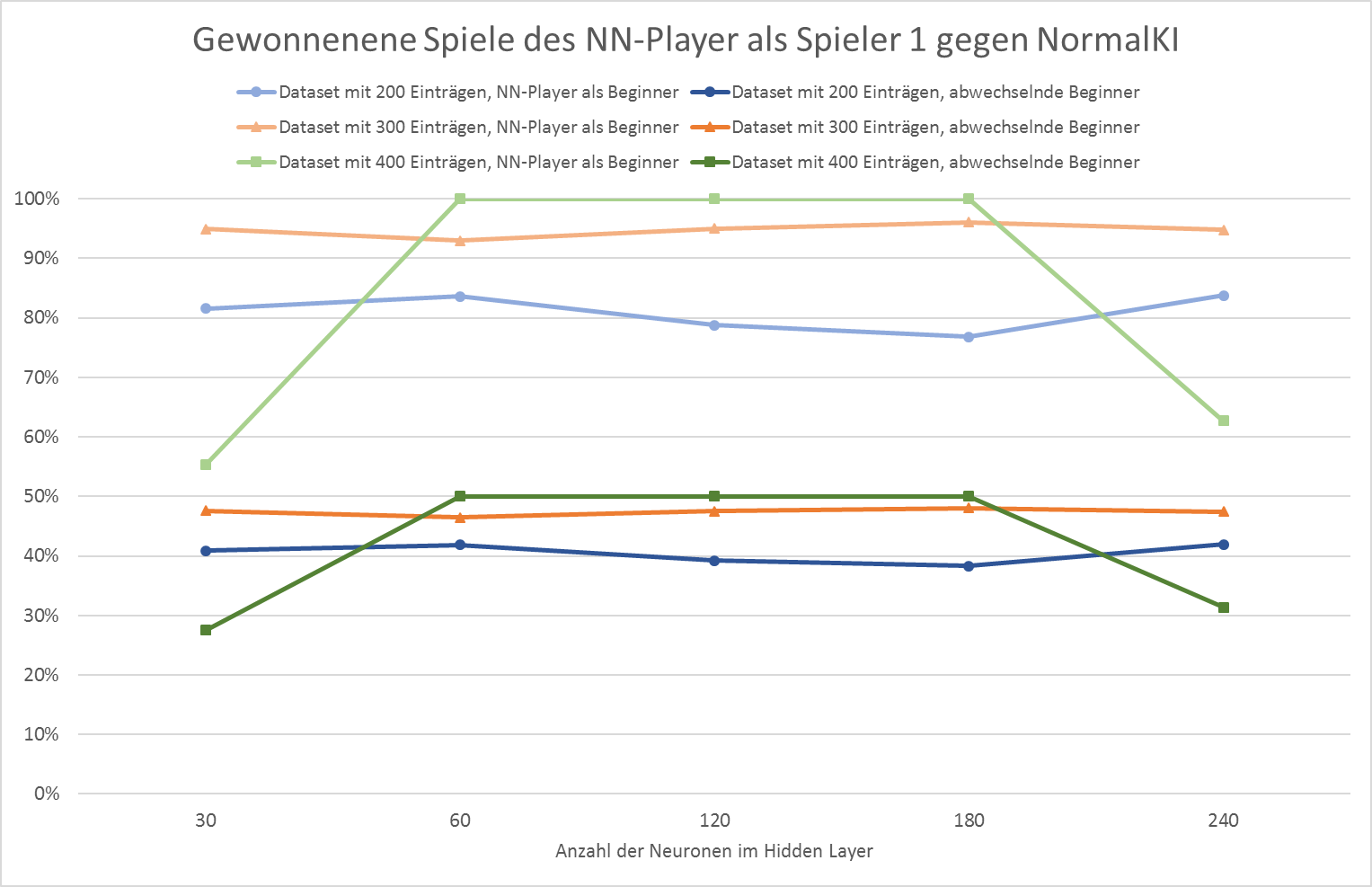


Abbildung 11

Abbildung 11 zeigt die Anzahl der im Mittel gewonnenen Spiele des *NNPlayer2* mit Neuronalen Netzen unterschiedlicher Neuronen-Anzahl im Hidden-Layer gegen die *NormalKI*. Die angewendeten Netze, wurden jeweils auf den Datasets von Spieler 1 trainiert.

Es ist erkennbar, dass sich die Kurven der NN-Player, die auf dem gleichen Dataset gelernt haben, aber unterschiedlich im Turnier beginnen, ähneln. Sehr eindeutig sieht man das auf den Kurven des Datasets mit 400 Einträgen: Die schlechtesten Ergebnisse liefert hier das Neuronale Netz mit 30 Neuronen im Hidden-Layer, während die Netze mit 60, 120 oder 180 Neuronen im Abwechselnd-Modus 50 % der Spiele gewinnt bzw. als Beginner der Spiele sogar 100 % der Spiele gewinnt, aber bei der Anzahl von 240 Neuronen wieder sinkt. Die Netze, die auf diesem Dataset mit 400 Einträgen gelernt haben und eine Neuronen-Anzahl von 60, 120 oder 180 aufweisen, können also vermutlich sehr gut einschätzen, wie sich die *NormalKI* als Spieler verhalten wird. Dass im Abwechselnd-Modus nur noch 50 % der Spiele gewonnen werden, statt 100 % als Beginner der Spiele, deutet daraufhin, dass das Netz alle Vorgehensweisen der *NormalKI* imitiert, also auch die des Verlierens. Was durchaus möglich ist, da in den Datasets sowohl gewonnene, verlorene als auch unentschiedene Spiele verzeichnet werden konnten.

Wenn man sich die Ergebnisse der Daten-Sets mit 200 und 300 Einträgen ansieht, so sind die Abweichungen der gewonnenen Spiele eines jeweiligen Daten-Sets im Vergleich zur Neuronen-Anzahl im Hidden-Layer meist nah beieinander und es sind nur leichte Schwankungen zu verzeichnen. Demnach ist ein Neuronales Netz, das gute Ergebnisse liefert, eines, welches auf dem Daten-Set mit 400 Einträgen gelernt hat und einer Neuronen-Anzahl von 60, 120 oder 180 im Hidden-Layer aufweist.

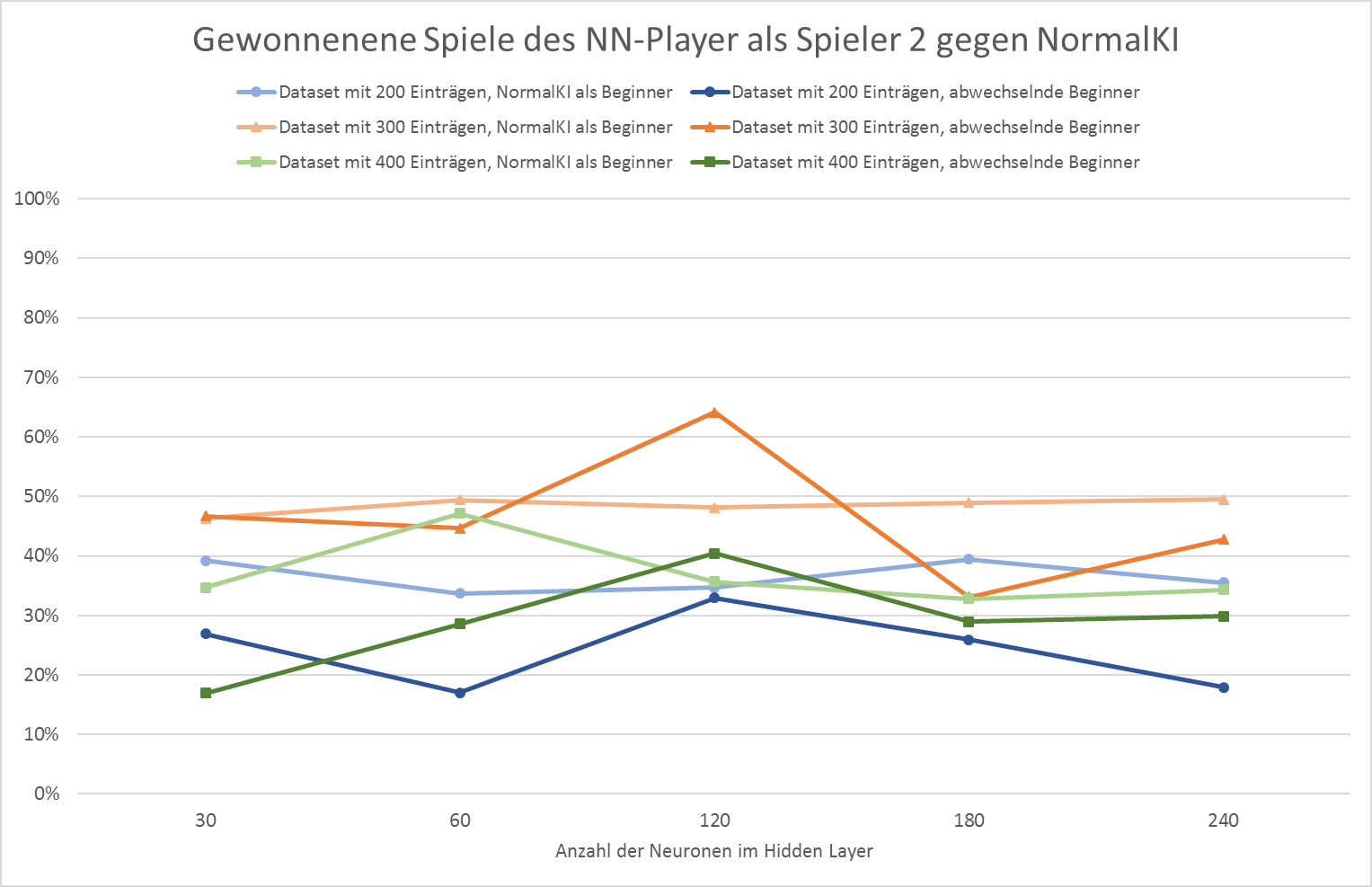


Abbildung 12

Abbildung 12 zeigt die Ergebnisse der im Mittel gewonnenen Spiele des *NNPlayer2* mit Neuronalen Netzen unterschiedlicher Neuronen-Anzahl im Hidden-Layer gegen die *NormalKI*, wobei hier die angewandten Netze, jeweils auf den Datasets von Spieler 2 gelernt haben. Zuerst fällt auf, dass die Ergebnisse insgesamt schlechter ausfallen als die Ergebnisse bei den Datasets mit Einträgen des 1. Spielers, welches die Vermutung stützt, dass der 1. Spieler (Beginner des Spiels) im Vorteil ist. Einzig das Neuronale Netz mit 120 Neuronen im Hidden-Layer, das auf einem Dataset mit 300 Einträgen des 2. Spielers gelernt hat, kann im Turnier eine Gewinnquote von 64 % im Abwechselnd-Modus aufweisen, während alle anderen Netze die 50 %-Schwelle nicht überschritten. Diese Beobachtung lässt vermuten, dass das Neuronale Netz aufgrund der Datasets des 2. Spielers wahrscheinlich hauptsächlich gelernt hat, auf eine Spielsituation zu reagieren und daher vermutlich eher die Strategie verfolgt die Gewinnchancen des Gegners zu verhindern. Um diese Vermutung zu stützen, müssten sich allerdings die Spielverläufe genauer angesehen werden.

Es ist außerdem zu beobachten, dass im Abwechselnd-Modus für die Neuronalen Netze, die auf den drei unterschiedlichen Datasets gelernt haben, eine Neuronen-Anzahl von 120 im Hidden-Layer im Test die besten Ergebnisse lieferten.

Spielt der *NNPlayer2* aber nur als 2. Spieler und die *NormalKI* beginnt im Turnier (Beginner-Modus für die *NormalKI*), so sind die Beobachtungen nicht mehr so eindeutig und schwanken stärker, was die Brauchbarkeit Neuronen-Anzahl im Hidden-Layer betrifft.

Im Test der Neuronalen Netze gab es ein paar Turnierverläufe, in denen es auch unentschiedene Spiele gab, allerdings waren hier nur die Spiele betroffen, von den Datasets mit 300 und 400 Einträgen, welche vom 2. Spieler aufgezeichnet wurden und hier waren maximal 6 % der Spiele in einem Turnier unentschieden und ist nicht sehr aussagekräftig, da es sein kann, dass es bei Generierung der anderen Datasets vielleicht nie zu unentschiedenen Spielen kam und das Netz diesen Spielverlauf vielleicht gar nicht kennt.

Wenn man sich ein paar Spielverläufe näher ansieht, fällt auf, dass der NNPlayer2 die Strategie der Zwickmühlen gelernt hat und dadurch ein Spiel für sich entscheiden kann. Die folgenden Spielverläufe beziehen sich auch einen NNPlayer2. Bei dem Neuronalen Netz handelt es sich um das, welches 240 Neuronen im Hidden-Layer hat und auf einem Daten-Set vom 1. Spieler mit 200 Einträgen gelernt hat. (*NNplayer2* = ■ *NormalKI*=X)

|  |  |  |
| --- | --- | --- |
| |0|0|0|0|0|  |0|0|0|0|0|  |0|0|0|0|0|  |0|0|■|X|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|0|0|■|0|  |0|0|■|X|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|0|0|■|0|  |0|0|■|X|X| |
| |0|0|0|0|0|  |0|0|0|0|0|  |0|0|0|■|■|  |0|0|■|X|X| | |0|0|0|0|0|  |0|0|0|0|0|  |0|0|X|■|■|  |0|0|■|X|X| | |0|0|0|0|0|  |0|0|0|0|■|  |0|0|X|■|■|  |0|0|■|X|X| |
| *Abbildung 13* | | |

In Abbildung 13 ist ein Spiel zu sehen, in dem der NNPlayer2 beginnt und innerhalb weniger Züge eine Zwickmühle baut.

|  |  |  |
| --- | --- | --- |
| |0|0|0|0|0|  |0|0|■|0|0|  |0|0|X|0|0|  |0|■|■|X|0| | |0|0|0|0|0|  |0|0|■|0|0|  |0|0|X|0|0|  |X|■|■|X|0| | |0|0|0|0|0|  |0|0|■|0|0|  |■|0|X|0|0|  |X|■|■|X|0| |
| |0|0|0|0|0|  |0|0|■|0|0|  |■|X|X|0|0|  |X|■|■|X|0| | |0|0|0|0|0|  |0|■|■|0|0|  |■|X|X|0|0|  |X|■|■|X|0| | |0|0|0|0|0|  |0|■|■|0|0|  |■|X|X|X|0|  |X|■|■|X|0| |
| *Abbildung 14* | | |

Abbildung 14 zeigt die letzten Züge eines Spielverlaufs des gleichen NN-Players. Der *NNPlayer2* verliert hier aufgrund einer Zwickmühle. Allerdings ist erkennbar, dass das Neuronale Netz die Zwickmühle – wenn auch zu spät - zu erkennen scheint (rot markiert), da er im letzten Zug die Diagonale (gelb markiert) blockiert.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |0|0|0|0|0|  |0|0|0|0|0|  |0|0|0|0|0|  |0|■|0|0|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|X|0|0|0|  |0|■|0|0|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|X|0|0|0|  |0|■|■|0|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|X|0|0|0|  |X|■|■|0|0| | |0|0|0|0|0|  |0|0|0|0|0|  |0|X|0|0|0|  |X|■|■|■|0| |
| *Abbildung 15* | | | | |

In Abbildung 15 wird ein Spielverlauf gezeigt, bei dem das Neuronale Netz mit 120 Neuronen im Hidden-Layer auf einem Daten-Set vom 2.Spieler mit 300 Einträgen gelernt hat. Auch hier wurde innerhalb weniger Züge eine Zwickmühle gebaut, die zum Sieg führte.

|  |  |  |
| --- | --- | --- |
| |0|0|0|0|0|  |0|X|0|0|0|  |0|X|■|0|0|  |0|■|X|0|0| | |0|■|0|0|0|  |0|X|0|0|0|  |0|X|■|0|0|  |0|■|X|0|0| | |0|■|0|0|0|  |0|X|X|0|0|  |0|X|■|0|0|  |0|■|X|0|0| |

|  |  |  |  |
| --- | --- | --- | --- |
| |0|■|0|0|0|  |0|X|X|0|0|  |0|X|■|0|0|  |■|■|X|0|0| | |0|■|X|0|0|  |0|X|X|0|0|  |0|X|■|0|0|  |■|■|X|0|0| | |0|■|X|0|0|  |0|X|X|0|0|  |■|X|■|0|0|  |■|■|X|0|0| | |0|■|X|0|0|  |X|X|X|0|0|  |■|X|■|0|0|  |■|■|X|0|0| |
| *Abbildung 16* | | | |

Abbildung 16 zeigt die letzten Züge eines Spielverlaufs des gleichen NN-Players, er auch in Abbildung 15 zu sehen ist. Es ist ersichtlich, dass das Neuronale Netz alle Gefahren (rot markiert) erkennt und sie innerhalb des nächsten Zuges verhindert (gelb markiert), aber leider aufgrund einer Zwickmühle des Gegners am Ende verliert.

Abschließend lässt sich also sagen, dass die Neuronalen Netze des *NNPlayer2* in der Lage zu sein scheinen, selbst Zwickmühlen zu bauen, aber auch auf mögliche Gefahren des Gegners zu reagieren, was ein strategisches Vorausdenken zeigt.

**NNPlayer2 6x7**

Da das Spiel auf einem 6x7 Feld und einer Gewinnbedingung von 4 Steinen deutlich komplexer wird, sind nicht die gleichen Ergebnisse wie auf dem 4x5 Feld zu erwarten gewesen. Trotzdem zeigt sich ein Lernerfolg durch Neuronalen Netzes und eine interessante Spielstrategie, die immer wieder zu gewonnenen Spielen für den NN-Player führten.

Um zu bestätigen, dass die KI trotz geringerer Gewinnquote eine deutliche Verbesserung aufweist, wird die Gewinnrate von einer zufälligen KI betrachtet: Diese gewinnt bei 10.000 Turnierspielen gerade einmal 138 also ca. 1 % der Spiele. Das Neuronale Netz ist wieder ein Multi-Layer Perzeptron und es wurden unterschiedliche Daten-Sets sowie unterschiedliche Parameter getestet. Die Anzahl der Input-Neuronen ergab 126, die Output-Neuronen 7 und die Hidden-Neuronen wurden während der Tests bei 252 belassen.

Es wurde mit unterschiedlich große Daten-Sets, welche mit der *generateDataSets*-Methode erstellt wurden, getestet:

1. 50 Trainingsspiele mit ca. 700 Einträgen
2. 100 Trainingsspiele mit ca. 1250 Einträgen
3. 150 Trainingsspiele mit ca. 2000 Einträgen
4. 500 Trainingsspiele mit ca. 6000 Einträgen

Dabei wurden für den maximalen Fehler Werte von 0,3 bis 0,01 verwendet. Eine Reduzierung auf einen Fehler unter 0,1 war allerdings nur bei einem kleineren Daten-Set möglich, da die Fehlerrate bei den größeren Sets davor bereits konvergiert hat. Auch Veränderungen an anderen Parametern, wie der Lernrate oder dem Momentum, konnten Erfolge verzeichnen. Bei 50 und 100 Trainingsspielen blieb die Gewinnquote bei ca. 10 % mit nur leichten Schwankungen bei unterschiedlichen Fehlerraten. Mit 150 Trainingsspielen (ca. 2000 Einträge) konnte das beste Ergebnis erzielt werden. Mit mehr Einträgen wurde das Ergebnis wieder schlechter. Das Ergebnis mit 100 Trainingsspielen lässt darauf schließen, dass ein zu geringer erlaubter Fehler zu einem Overfitting führt und der NN-Player daher schlecht spielt. Dies bestätigt das Ergebnis mit 150 Trainingsspielen. Bei einem Fehler von 0,3 gewann der Spieler ca. 25 % der Spiele und 10 % gingen unentschieden aus. Im Vergleich zu den anderen Trainingsversuchen und dem Spielen gegen eine zufällig spielende KI ist das eine deutliche Mehrleistung.

Im Folgenden soll die Spielweise des NN-Players mit dem eben beschriebenen, besten neuronalen Netz betrachtet werden (*NNPlayer* = ■ *NormalKI*=X)

|  |  |  |
| --- | --- | --- |
| Spiel 1 | Spiel 2 | Spiel 3 |

|  |  |  |
| --- | --- | --- |
| |0|0|■|X|X|0|0|  |0|0|X|■|X|0|0|  |0|0|■|■|■|0|0|  |0|0|■|■|X|0|0|  |0|X|■|X|■|■|0|  |■|X|X|■|X|X|0|  |0|0|■|X|X|0|0|  |0|0|X|■|X|0|0|  |0|0|■|■|■|0|0|  |0|0|■|■|X|X|0|  |0|X|■|X|■|■|0|  |■|X|X|■|X|X|0|  |0|0|■|X|X|0|0|  |0|0|X|■|X|0|0|  |0|0|■|■|■|■|0|  |0|0|■|■|X|X|0|  |0|X|■|X|■|■|0|  |■|X|X|■|X|X|0| | |0|0|X|0|■|0|0|  |0|0|■|X|X|0|0|  |0|0|X|■|■|0|0|  |0|0|■|■|■|0|0|  |0|0|■|X|X|0|0|  |0|X|X|■|X|■|0|  |0|0|X|0|■|0|0|  |0|0|■|X|X|0|0|  |0|0|X|■|■|0|0|  |0|0|■|■|■|0|0|  |0|0|■|X|X|X|0|  |0|X|X|■|X|■|0|  |0|0|X|0|■|0|0|  |0|0|■|X|X|0|0|  |0|0|X|■|■|0|0|  |0|0|■|■|■|■|0|  |0|0|■|X|X|X|0|  |0|X|X|■|X|■|0| | |0|0|0|0|0|0|0|  |0|0|X|X|0|0|0|  |0|0|■|■|0|0|0|  |0|0|■|X|■|0|0|  |0|■|■|■|X|0|0|  |0|X|X|■|X|0|0|  |0|0|0|0|0|0|0|  |0|0|X|X|0|0|0|  |0|0|■|■|0|0|0|  |0|0|■|X|■|0|0|  |0|■|■|■|X|0|0|  |X|X|X|■|X|0|0|  |0|0|0|0|0|0|0|  |0|0|X|X|0|0|0|  |0|0|■|■|0|0|0|  |0|0|■|X|■|0|0|  |■|■|■|■|X|0|0|  |X|X|X|■|X|0|0| |

*Abbildung 17*

Diese drei Spiele zeigen, dass das Neuronale Netz gelernt hat, Spielsituation zu erschaffen, in denen der Gegner in eine Zwickmühle gerät. Diese Situationen treten in nahezu allen Spielen auf, die der NN-Player gewinnt.

|  |  |  |
| --- | --- | --- |
| Spiel 1 | Spiel 2 | Spiel 3 |
| |X|■|■|X|X|0|X|  |■|■|X|■|X|0|■|  |X|X|■|■|X|X|X|  |■|■|■|X|■|■|■|  |X|X|■|■|X|X|X|  |■|X|X|■|X|■|■|  |X|■|■|X|X|0|X|  |■|■|X|■|X|■|■|  |X|X|■|■|X|X|X|  |■|■|■|X|■|■|■|  |X|X|■|■|X|X|X|  |■|X|X|■|X|■|■|  |X|■|■|X|X|X|X|  |■|■|X|■|X|■|■|  |X|X|■|■|X|X|X|  |■|■|■|X|■|■|■|  |X|X|■|■|X|X|X|  |■|X|X|■|X|■|■| | |0|0|■|X|0|0|0|  |0|0|X|■|0|0|0|  |0|0|■|X|0|0|0|  |0|0|■|■|X|0|0|  |0|0|■|X|■|0|0|  |0|0|X|■|X|X|0|  |0|0|■|X|0|0|0|  |0|0|X|■|0|0|0|  |0|0|■|X|■|0|0|  |0|0|■|■|X|0|0|  |0|0|■|X|■|0|0|  |0|0|X|■|X|X|0|  |0|0|■|X|0|0|0|  |0|0|X|■|0|0|0|  |0|0|■|X|■|0|0|  |0|0|■|■|X|0|0|  |0|0|■|X|■|X|0|  |0|0|X|■|X|X|0| | |0|0|X|X|■|0|0|  |0|0|■|■|■|0|0|  |0|0|X|X|X|0|0|  |0|0|■|■|■|0|0|  |0|0|■|X|X|0|X|  |0|X|■|■|X|X|■|  |0|0|X|X|■|0|0|  |0|0|■|■|■|0|0|  |0|0|X|X|X|0|0|  |0|0|■|■|■|0|■|  |0|0|■|X|X|0|X|  |0|X|■|■|X|X|■|  |0|0|X|X|■|0|0|  |0|0|■|■|■|0|0|  |0|0|X|X|X|0|0|  |0|0|■|■|■|0|■|  |0|0|■|X|X|X|X|  |0|X|■|■|X|X|■| |

*Abbildung 18*

Die die in Abbildung 18 zu sehenden Spiele stellen häufig wiederzufindende Spielsituationen aus den Turnierspielen dar. Wenn die *NormalKI* gewinnt, dann immer über eine einfache Kombination, die der NN-Player übersieht. Zwickmühlen werden von der *NormalKI* nicht erzeugt. Weiterhin lässt sich beobachten, dass die Spiele häufig sehr lange dauern. Die durchschnittliche Anzahl an Zügen pro Spiel liegt bei ca. 29. D.h. Spielzustände wie der aus Spiel 1 oder ähnliche, die unentschieden ausgehen, sind häufig in den Spielen zu finden.

Insgesamt kann man das Potenzial der Neuronalen Netze auch bei einem Spielfeld der Größe 6x7 erkennen. Es hätte mehr Zeit gebraucht, um noch mehr Einstellungen zu testen. Insbesondere bei den Daten-Sets würde es durchaus Verbesserungsmöglichkeiten geben. Da die Daten-Sets automatisch generiert werden, schwankt die Gewinnquote des NN-Players erheblich. Mit einem manuell erstellten Daten-Set oder einer anderen Art Daten-Sets zu erstellen, könnten weitere Verbesserungen möglich sein. Eine Möglichkeit wäre, bei der Erstellung darauf zu achten, wie gut die einzelnen Einträge in dem Daten-Set sind.

**Q-Player gegen NN-Player 4x5**

In einer letzten Testreihe traten *QPlayer2* und NNPlayer2 auf dem 4x5 Feld gegen einander an. Für beide Spieler wurden die besten Versionen aus den vorhergegangenen Tests genommen. Der Q-Player erhielt dafür seine Datenbank mit 32.000 Trainingsspielen, bei der die Gewinnquote gegen die NormalKI2 bei ca. 80% lag. Der NNPlayer2 wurde nur für die Züge des zweiten Spielers trainiert und hatte gegen die *NormalKI* eine Gewinnquote von ca. 64 %. Es befanden sich 300 Einträge im gelernten Daten-Set und das Neuronale Netz weist 120 Neuronen im Hidden-Layer auf.

Überraschender Weise schlug der *NNPlayer2* den *QPlayer2* deutlich. In fünf Turnieren mit 10.000 Spiele, bei denen abwechselnd begonnen wurde, gewann der *QPlayer2* durchschnittlich nur 23,73%. Auch in einem zweiten Durchlauf, in dem der *QPlayer2* in jeder Runde anfangen durfte, lag seine Gewinnquote nur bei 27,76%.

Auffällig war, dass der *QPlayer2* in den Turnieren durchschnittlich 72619,6 bzw. 36004,6 Zustände nicht kannte. Es liegt daher nah, dass sich der *Q-Player* im Training zu sehr auf die NormalKI2 eingestellt hat und daher nicht mit dem *NNPlayer2* umzugehen weiß.

Aufgrund dieser Vermutung wurden dieselben Turniere noch einmal mit einer größeren Datenbank für den *QPlayer2* durchgeführt. Doch trotz der nun vierfachen Menge an Trainingsdurchläufen und ca. 1,7-fachen Menge an Datenbankelementen bliebt die Performance ähnlich schlecht. Der *QPlayer2* verbesserte sich beim abwechselnden Spielbeginn um nur 0,312%. Wenn er in jedem Spiel begann, lag die Verbesserung gerade einmal bei 0,072%.

Bei genauer Betrachtung der Spieleverläufe ist festzustellen, dass auch der NN-Player nicht optimal auf die Spielsituationen eingehen kann. Das folgende Beispiel soll dies veranschaulichen.  
( QPlayer2 = ■ NNPlayer2 = X )

|  |  |  |  |
| --- | --- | --- | --- |
| |0|0|0|X|0|  |0|0|0|■|0|  |0|0|0|X|0|  |0|X|0|■|■| | |0|0|0|X|0|  |0|0|0|■|0|  |0|0|0|X|X|  |0|X|0|■|■| | |0|0|0|X|0|  |0|0|0|■|0|  |0|■|0|X|X|  |0|X|0|■|■| | |0|0|0|X|0|  |0|0|0|■|X|  |0|■|0|X|X|  |0|X|0|■|■| |
| |0|0|0|X|0|  |0|0|0|■|X|  |0|■|0|X|X|  |■|X|0|■|■| | |0|0|0|X|0|  |0|X|0|■|X|  |0|■|0|X|X|  |■|X|0|■|■| | |0|0|0|X|0|  |0|X|0|■|X|  |■|■|0|X|X|  |■|X|0|■|■| | |0|0|0|X|0|  |X|X|0|■|X|  |■|■|0|X|X|  |■|X|0|■|■| |
| |0|0|0|X|■|  |X|X|0|■|X|  |■|■|0|X|X|  |■|X|0|■|■| | |0|X|0|X|■|  |X|X|0|■|X|  |■|■|0|X|X|  |■|X|0|■|■| | |■|X|0|X|■|  |X|X|0|■|X|  |■|■|0|X|X|  |■|X|0|■|■| | |■|X|0|X|■|  |X|X|0|■|X|  |■|■|0|X|X|  |■|X|X|■|■| |

*Abbildung 19*

In Abbildung 19 ist deutlich zu erkennen, dass keiner der Spieler mit der Spielsituation vertraut ist, bzw. eine gute Strategie für die vorliegenden Spielzustände kennt. Schon recht früh im Spielverlauf ergibt sich für den Q-Player die Möglichkeit drei Steine in Reihe zu positionieren. In den folgenden Zügen versucht der NN-Player dies nicht zu verhindern, der Q-Player nutzt seine Chance jedoch auch nicht. Im weiteren Verlauf positioniert der NN-Player seine Steine so, dass er selbst über das gleiche Feld eine Sieg-Chance hat wie die Q-Player. Beide Spieler müssten in die mittlere Spalte werfen, um den Sieg des anderen zu verhindern und den eigenen zu erlangen. Keiner von beiden nutzt diese Möglichkeit. Erst als alle anderen Spalten voll sind, wirft der NN-Player in die Mitte und gewinnt.

## Fazit und Ausblick

Die Ergebnisse dieses Business Analytics Projekts zeigen, dass es mit der Methode des Q-Learnings und auch mit der Methode von Neuronalen Netzen möglich ist, eine selbst lernende Künstliche Intelligenz für das Spiel „Vier-Gewinnt“ zu implementieren, doch gibt es noch weitere Möglichkeiten dieses Projekt in Zukunft fortzuführen bzw. zu erweitern, diese werden im Folgenden dargestellt.

**Q-Player**

Der Q-Player erzielt auf einem 4x5 Feld hervorragende Ergebnisse zur *NormalKI,* trotzdem sollte ein Weg gefunden werden, um mit dem Overfitting zur *NormalKI* umzugehen. Die zweite Frage, die bleibt ist, warum er ab einem gewissen Grad wieder schlechter wird. Hierzu sollte die Berechnung der Werte für die Zustände neu betrachtet werden. Evtl. genügt es schon auf die Durchschnittsberechnung zu verzichten und nur den maximalen Wert der nächsten Zustände zu betrachten, nachdem für den Gegner vom für ihn bestmöglichen Zug ausgegangen wird. Ein anderer Ansatz ist es nicht mehr davon auszugehen, dass der Gegner den bestmöglichen Zug macht, sondern stattdessen an dieser Stelle einen Mittelwert zu bilden. Beide Alternativen müssten experimentell erforscht werden.

Auch auf dem 6x7 Spielfeld kann der Q-Player von einer Verbesserung an dieser Stelle profitieren. Zunächst muss jedoch sichergestellt werden, dass er überhaupt alle für ihn wichtigen Spielzustände erforschen kann. Da die gewählte Implementierung der Datenbank hier an ihre Grenzen stößt, liegt hier Verbesserungspotenzial vor, um den Q-Player auch auf einem 6x7 Feld zu trainieren. Die Datenbank wird sehr groß und ab einer gewissen Größe reicht der Arbeitsspeicher nicht mehr aus. Dann wird die Datenbank zu Teilen auf die Festplatte geschrieben und das Abfragen von Spielzuständen dauert sehr lange. Ein Ansatz zur Komprimierung der Datenbank scheint lohnenswert. Der Schlüssel der HashMap, die intern verwendet wird, ist ein 2-dimensionales Int-Array. Dies ist vom Speicherverbrauch nicht sehr effizient, da die Anzahl der Zustände in der Datenbank sehr schnell sehr groß wird. Bei 3.200.000 Trainings gibt es 21.055.352 Einträge in der Datenbank. Einen kleineren Schlüssel zu verwenden könnte den Speicherverbrauch reduzieren.

Allerdings kann es sein, dass auch so nicht alle Zustände in einer Datenbank untergebracht werden können. Wie oben erwähnt beschäftigt sich Olszowka in seiner Arbeit mit der Problematik rund um die Anzahl der Zustände bei einem 6x7 Spielfeld. Er geht sogar davon aus, dass es nicht einmal möglich ist, die KI alle Zustände erforschen zu lassen. Diese Aussage basiert jedoch auf der Annahme, dass es etwa Zustände gibt. [vgl. 6] Tatsächlich sollte die Anzahl bedeutend kleiner sein, da die Steine nicht beliebig auf dem Feld verteilt werden dürfen. Zu beachten sind u. a. folgende Punkte:

* Die Spieler dürfen nur abwechselnd werfen, d. h. die Anzahl der Steine von Spieler 2 und Spieler 2 unterscheiden sich maximal um eins.
* Steine könne nicht „schweben“. Sie fallen immer so weit nach unten, wie es ihnen möglich ist.
* Wenn vier Steine einer Farbe nebeneinander liegen, ist das Spiel beendet, d.h. es werden keine weiteren Steine mehr platziert.

Trotzdem könnte es sich lohnen, sich mit Optionen zur Zustandsreduzierung zu befassen. Olszowka schlägt die Verwendung von Feature-Vektoren vor. Dabei wird nicht mehr jeder Spielzustand einzeln gespeichert. Stattdessen erhält der der Q-Player einen Vektor mit Eigenschaften, die im jeweiligen Zustand enthalten sind. Wichtig ist aber, dass nicht zu viele Informationen verloren gehen. [vgl. 6]

**NN-Player**

Den *NNPlayer2* betreffend, wäre es von Bedeutung, sich die Daten-Sets genauer anzuschauen. Zurzeit werden sie durch zufällige Spielzüge der *NormalKI* generiert und man hat bisher keinen Einfluss darauf, bestimmte Spielzüge auszusortieren, d. h. es ist möglich, dass auch Spielzüge gespeichert werden, die zum Verlieren führen. Daher wäre es sinnvoll, nur die Spielzüge zu speichern, die zum Sieg führen.

Zudem sollte nach einer Möglichkeit gesucht werden, den NN-Player weniger auf Spielzüge der *NormalKI* auszurichten. Auch wenn der NNPlayer2 weiniger vom Overfitting betroffen zu sein scheint als der *QPlayer2,* zeigen die Spielverläufe, dass er mit einigen Situationen nicht umgehen kann.

Zuletzt kann die *turn*-Methode im *NNPlayer2* umgestaltet werden, damit hier mit zwei Neuronalen Netzen gearbeitet werden kann; eines für die Spiele des 1. Spielers und eines für Spiele des 2. Spielers. Je nach Situation im Turnier, sollte dann das passende Netz ausgewählt werden, um so die Gewinnchancen zu erhöhen. Dies müsste allerdings auch experimentell getestet werden.

# Verweise

[1] Silver, D. et al.: Mastering the game of Go with deep neural networks and tree

search. Nature (529), 28.01.2016, S. 484–489.

[2] Allis, L. V.: Searching for Solutions in Games and Artificial Intelligence (Ph.D. thesis).

University of Limburg, Maastricht, 1994.

[3] Edelkamp, S., Kissmann, P.: Symbolic Classification of General Two-Player Games.

Technical report, Technische Universität Dortmund, 2008.

[4] <http://www.java-online.ch/gamegrid/gamegridEnglish/index.php?inhalt_mitte=gittergames/fourInARow.inc.php> [06.09.2016]

[5] Watkins, C. J. C. H., Dayan, P.: Q-learning. Machine Learning (8), 1992, S. 279–292.

[6] Olszowka, D.: Reinforcement-Learning in der Domäne von PL-1. AW-1 Ausarbeitung,

HAW Hamburg.

[7] Șimșek, Ö., Barto, A. G.: Skill characterization based on betweenness. NIPS´08, 2008, S. 1497-1504.

[8] <http://mnemstudio.org/path-finding-q-learning-tutorial.htm> [06.09.2016]

[9] Tokic, M., Palm, G.: Value-Difference Based Exploration: Adaptive Control between Epsilon-Greedy and Softmax. In J. Bach & S. Edelkamp (Eds.), Proceedings of the 34th annual German conference on advances in artificial intelligence, Berlin: Springer-Verlag, 2011, S. 335-346.

[10] MacKay, D. J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge

University Press, 2003, ISBN 9780521642989.

[11] <ftp://ftp.sas.com/pub/neural/FAQ2.html#A_act> [06.09.2016]

[12] Rumelhart, D. E., Hinton, G. E., Williams, R. J.: Learning representations by back-

propagating errors. Nature 323 (6088), 08.10.1986, S. 533–536.

[13] Schneider, M.O., Rosa, J.L.G.: Neural Connect Four – A Connectionist Approach to

the Game. Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN’02), 2002.

[14] Chellapilla, K., Fogel, D. B.: Evolution, neural networks, games, and intelligence.

Proceedings \_. of the IEEE vol. 87 (9), 1999, S. 1471-1496.

[15] <http://neuroph.sourceforge.net/Getting%20Started%20with%20Neuroph%202.7.pdf> [06.09.2016]

[16] <http://neuroph.sourceforge.net/tutorials/MultiLayerPerceptron.html> [06.09.2016]