**Q-Learning 4x5**

Der Q-Player wurde mit einem *epsilon* von 0,2 und einem *alpha* von 0,05 schrittweise in insgesamt 128.000 Spielen trainiert. Die Anschließenden Test wurden in fünf Turnieren mit je 10.000 Spielen durchgeführt.

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. 1]

Abbildung 1

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. 2]

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%.

Abbildung 2

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 ehr verhindern kann. Die folgenden drei Spielausschnitte stammen vom Q-Player mit 32.000 trainingsspielen und sollen dieses Verhalten veranschaulichen. (Q-Player = ■ 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|

Der Q-Player 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 Q-Player die Gewinnmöglichkeit über die Diagonale gibt.

|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|■|■|■|

In diesem Spiel positioniert der Q-Player seine Steine so, dass er sowohl in der untersten als auch in der darüber liegenden Zeile eine Dreierreihe vervollständigen kann.

|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|

Auch hier hat der q-Player im zweiten gezeigten zustand zwei Möglichkeiten zu gewinnen. Zum einen in die vierte 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.

**Q-Learning 6x7**

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

Abbildung 3

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

Ein weiteres Training auf 4,8Mio bzw. 6,4Mio Trainingsspiele wurde zwar versucht, jedoch abgebrochen, da der Trainingsprozess zunehmend langsamer wurde. Zuletzt lag die Anzahl der Spiele durchgeführten bei ca. 200 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 4

Trotz der geringen Erfolgsquote nach 3.2MIo Lerndurchläufen lassen sich einige geschickte Züge in den Turnierspielen finden. Genauso wie auf dem 4x5 Feld konstruiert der Q-Player 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| |

Das erste Beispiel ist eine Einfach Form der Zwickmühle, bei der der Q-Player 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| |

Im zweiten Beispiel ist eine Komplexere Strategie zu sehen. Der Q-Player 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 Q-Player verhindert, gleichzeitig ist es ihm nun aber möglich eine Vierer-Reihe in der Horizontalen zu vervollständigen.