**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 würden mit Turnieren á 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. [s. 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. Was nicht zu erwarten war, ist dass, jeder Versuch sie weiter zu trainieren, auch mit einer sehr geringen Anzahl von zusätzlichen Spielen, zu einer Verschlechterung der Spielperformance führte. [s. 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 Vorausdenden 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=1)

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

|0|0|2|0|0| |0|0|2|0|0| |0|0|2|0|0| |1|0|2|0|0|

|0|0|1|0|0| |0|1|1|0|0| |2|1|1|0|0| |2|1|1|0|0|

|1|2|1|2|0| |1|2|1|2|0| |1|2|1|2|0| |1|2|1|2|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|2|2|0|0| |0|2|2|0|0| |0|2|2|0|0| |0|2|2|0|0|

|0|1|1|0|0| |0|1|1|0|0| |0|1|1|2|0| |0|1|1|2|0|

|0|2|1|0|0| |0|2|1|1|0| |0|2|1|1|0| |0|2|1|1|1|

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|2|0|0|0| |0|2|0|0|0| |0|2|0|0|0| |0|2|0|1|0|

|0|2|1|0|0| |0|2|1|0|0| |0|2|1|2|0| |0|2|1|2|0|

|0|1|2|0|0| |0|1|2|1|0| |0|1|2|1|0| |0|1|2|1|0|

|2|1|2|1|0| |2|1|2|1|0| |2|1|2|1|0| |2|1|2|1|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.