Reinforcement Learning

Projektbeschreibung

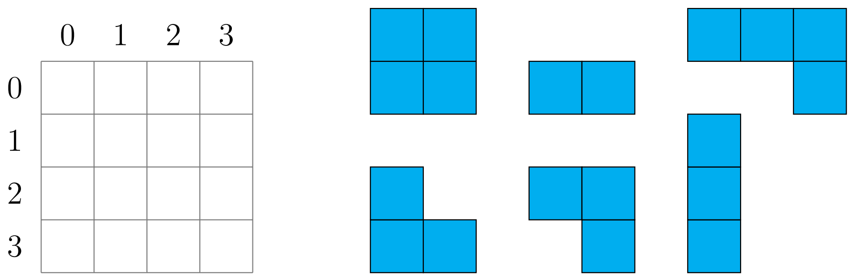
Patrizia Schalk
patrizia.schalk@student.uni-augsburg.de

Manuel Richter
manuel.david.richter@uni-augsburg.de

16. Januar 2018

1 Problembeschreibung

Gegeben sei ein Spielfeld der Größe x, sowie n Bausteine beliebiger Größe. Die Aufgabe besteht darin, m der n Bausteine so auf dem Spielfeld zu platzieren, dass möglichst wenige Felder frei bleiben.



Die Bausteine dürfen dabei weder gedreht noch gespiegelt werden - sie müssen exakt so verwendet werden, wie sie neben dem Spielfeld liegen. Ferner dürfen sich zwei Bausteine nicht überlappen oder über den Rand des Spielfeldes ragen.

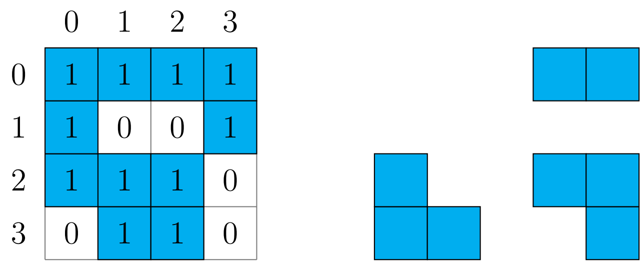
2 Umsetzung durch Reinforcement Learning

Das oben genannte Problem soll mittels Q-Learning mit den im Folgenden beschriebenen Parametern gelöst werden.

2.1 Zustandsraum

Ein Spielfeld der Größe n setzt sich zusammen aus n^2 Feldern, von denen zu einem beliebigen Zeitpunkt des Spiels $m \in \{0, ..., (n-1)^2\}$ durch einen Stein besetzt und $n^2 - m$ unbesetzt sind.
Sei $f : \{0, ..., n-1\} \times \{0, ..., n-1\} \to \{0, 1\}$ mit:

$$f(x,y) = \begin{cases} 1 & \text{falls } (x,y) \text{ durch einen Spielstein besetzt} \\ 0 & \text{falls } (x,y) \text{ unbesetzt} \end{cases}$$



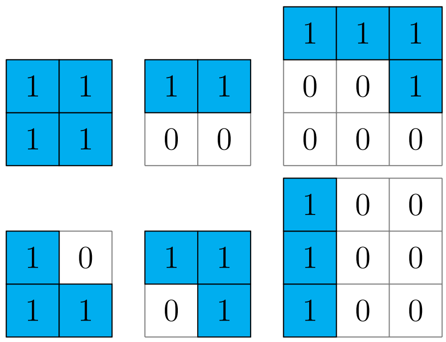
Mit dieser Darstellung des Spielfeldes lässt sich der aktuelle Zustand des Spielfelds durch die Binärzahl

$$(f(0,0), ..., f(0,n-1), f(1,0), ..., f(1,n-1), ..., f(n-1,0), ..., f(n-1,n-1))_2$$

darstellen. Diese Binärzahl hat genau $(n-1)^2$ Bits und kann damit in einer ganzen Zahl $z \in \{-2^{n-1}, ..., 0, ..., 2^{n-1}\}$ gespeichert werden.

Mit der Einschränkung der Zahlendarstellung durch Rechenmaschinen sind wir dadurch nicht in der Lage, Probleminstanzen der Größe > 30 zu lösen, wenn wir diesen Zustand in einem Integer speichern. Speichern wir den Zustand in einem Long, so können wir Probleminstanzen bis zur Größe 62 lösen.

Neben der Darstellung des Spielfeldes benötigen wir zusätzlich die Bausteine, welche dem Agenten noch zur Verfügung stehen. Da wir nicht ausschließen, dass sich die Form zweier Bausteine gleicht, verwenden wir dafür eine Multimenge μ. Die Darstellung der Bausteine erfolgt analog zum Spielfeld über eine Binärzahl. Zu diesem Zweck werden die Spielsteine wie dargestellt mit einer Maske erweitert:



Damit erhalten wir für die Darstellung eines jeden Bausteins ebenfalls eine ganze Zahl und können unsere Multimenge μ mit diesen Darstellungen füllen.

$\Rightarrow$ Der Zustandsraum $\mathcal{Z}$ setzt sich zusammen aus $S \subseteq \mathbb{Z}$ und μ.
$\Rightarrow \mathcal{Z} = S \times \mu$
Mit einer Anzahl c von Spielsteinen gilt $|\mathcal{Z}| < |S| \cdot |\mu| < 2^n \cdot c$.

2.2 Aktionen

Der Agent hat jeweils die Möglichkeit, einen Spielstein aus der Menge μ auf das Spielfeld zu setzen. Damit spannt sich also der Aktionsraum auf zu $\mathcal{A} = \{0, ..., n-1\} \times \{0, ..., n-1\} \times \mu$.
Mit den Konventionen aus dem vorherigen Abschnitt gilt damit:
$|\mathcal{A}| \leq c \cdot (n-1)^2$

2.3 Ende einer Episode

Die Episode endet, wenn der Agent versucht, einen illegalen Spielzug durchzuführen.