

Das Münzspiel

Ein Beispiel für verstärkendes Lernen

Informatikunterricht WP II, Unterrichtseinheit KI

Verstärkendes Lernen;

am Beispiel eines Zwei-Personen-Spiels

Stellen Sie sich vor, Sie treten in einem Zwei-Personen-Spiel gegen eine andere Person an. In diesem Spiel bauen Sie abwechselnd einen Zug zu machen, wobei Ihnen die jeweils erlaubten Züge angegeben werden. Irgendwann endet das Spiel und Sie erfahren, ob Sie gewonnen oder verloren haben.

So könnte es Ihnen ergehen, wenn Sie als Neuling z.B. Schach oder GO spielen sollen.

Wie können Sie lernen, welcher Zug in einer bestimmten Situation sinnvoll, gewinnversprechend oder schlecht ist?

Die Idee besteht darin, mehrere (viele) Spiele zu spielen. Immer dann, wenn Sie verlieren, werden gewissen Züge als "schlecht" markiert, während Züge, die zu Ihrem Gewinn geführt haben, für Sie weiterhin machbar bleiben.

Am Beispiel eines sehr einfachen Spiels soll das Vorgehen erläutert werden.

Das Spiel, die Regeln und eine Beispielrunde

Auf dem Tisch liegen einige Münzen. Zwei Personen (hier Alice und Bob genannt) spielen gegeneinander.

Sie nehmen abwechselnd eine oder zwei Münzen¹ weg und legen sie beiseite. Derjenige Spieler, der die letzte bzw. die letzten beiden Münzen wegnimmt², hat das Spiel verloren.

Hier gehen wir davon aus, dass zu Beginn einer Runde 5 Münzen auf dem Tisch liegen.

¹ In einer Spielvariante können auch mehr als zwei Münzen vom Tisch genommen werden.

² Eine andere Regel ist möglich, bei der derjenige Spieler gewinnt, der die letzte Münze bzw. die letzten Münzen nimmt.

Ein möglicher Spielablauf könnte wie folgt aussehen:

1. Es liegen 5 Münzen auf dem Tisch:



2. Alice beginnt und nimmt eine Münze; es liegen noch 4 Münzen auf dem Tisch:



3. Bob nimmt zwei Münzen, so dass jetzt noch zwei Münzen übrig bleiben:



4. Alice nimmt eine Münze; es bleibt eine Münze liegen:



5. Bob muss die letzte Münze nehmen und verliert.
6. Es könnte jetzt eine neue Runde beginnen, in der Bob den ersten Zug machen darf.

Bob möchte lernen!





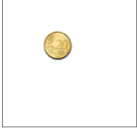
Wie jeder Spieler möchte Bob natürlich das Spiel gewinnen; doch er hat keine Ahnung, welchen Zug er in einer Spielsituation machen soll. Er erfährt nur am Ende einer Runde, ob er gewonnen oder verloren hat. Insbesondere hat man ihm nicht verraten, dass er verliert, wenn er die letzten Münze vom Tisch nimmt.

Er nimmt sich vor, mehrere Runden gegen Alice zu spielen und dabei zu lernen, welche Züge mehr oder weniger erfolgversprechend sind.

Die Idee

Jedesmal, wenn Bob am Zug ist, wird er sich diesen Zug merken.

Dazu legt Bob sich eine Tabelle an:





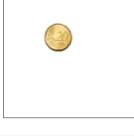
Zustand	geeignete bzw. sinnvolle Züge	Gemerakter Zug
		
		
		
		
		

Diese Tabelle besteht (in diesem Beispiel) aus fünf Zeilen und drei Spalten:

- In der linken Spalte findet man den **Spielzustand**: *Wieviele Münzen liegen momentan auf den Tisch?*
- In den Feldern der mittleren Spalte findet einige Zugmöglichkeiten, die in dem jeweiligen Zustand möglich, sinnvoll, vielversprechend, ... sind. Die Zugmöglichkeiten, die in diesen Feldern zu finden sind - so die Idee von Bob - werden sich im Verlauf des Lernprozesses ändern: es werden weniger werden! Denn ein Zug, der dazu führt, dass Bob verliert, wird dort nicht mehr erscheinen.

- Immer dann, wenn Bob in einem bestimmten Spielzustand einen Zug macht, merkt er sich diesen Zug in dem zugehörigen Feld in der rechten Spalte.

Die Tabelle hat Zu Beginn des Lernvorgangs folgendes Aussehen:

Zustand	geeignete bzw. sinnvolle Züge	Gemerakter Zug
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm eine Münze</div>	

Er hat auf die Felder in der mittleren Spalte Marken gelegt, die einen Zug beschreiben.

Im Laufe der Zeit “verschwinden” einige dieser Marken. Die dort (noch) vorhandenen Marken (*Nimm eine Münze* bzw. *Nimm zwei Münzen*) kann man folgendermaßen interpretieren:

- Führt Bob diesen Zug aus, könnte er ggf. noch gewinnen. Sicher ist das jedoch nicht.

Oder anders ausgedrückt:

- Führt Bob einen Zug aus, der dort nicht (mehr) zu finden ist, kann er das Spiel aus eigener Kraft nicht mehr gewinnen (also falls Alice optimal spielt). Er muss also hoffen, dass Alice im weiteren Verlauf einen Fehler macht.

Lernstrategie, die Idee

Das Ziel des Lernvorgangs besteht nun darin, dass irgendwann alle Züge, die in der mittleren Spalte zu finden sind, für Bob garantiert zum Sieg führen, unabhängig davon, welche Züge Alice macht.

Ist Bob am Zug, gibt es zwei Möglichkeiten:

1. Das zu der jeweiligen Situation gehörige Feld in der mittleren Spalte ist leer, es liegt hier keine Marke mehr.
 - Das bedeutet, dass es keinen Zug gibt, der einen Gewinn garantiert. Der letzte Zug von Bob (sofern er vorher schon einen Zug gemacht hat), war also nicht wirklich gut. Hat er sich diesen Zug mit der entsprechenden Marke im zugehörigen rechten Feld gemerkt, kann er diese Marke als *ungeeignet* aus dem Spiel entfernen.
 - Er wird dann irgendeinen Zufallszug machen, in der Hoffnung, dass Alice nicht optimal weiter spielt. Jedoch muss er sich diesen Zufallszug nicht merken.
2. Es gibt in dem zugehörigen mittleren Feld noch einen oder mehrere Marken (also Zugmöglichkeiten) zur Auswahl. das sind also Züge, die noch nicht als *ungeeignet* erkannt sind. Diese Züge lassen also noch auf einen Sieg hoffen.
 - Der letzte Zug von Bob (sofern er vorher schon einen Zug gemacht hat) war also nicht schlecht, so dass er diese Marke aus dem zugehörigen rechten Feld wieder zurück in die mittlere Spalte schiebt. Stattdessen wählt er (zufällig) einen Zug aus, merkt sich diesen mit der entsprechenden Marke im rechten Feld und führt ihn aus.

Lernstrategie, die praktische Umsetzung

Immer dann, wenn Bob einen Zug machen soll, geht er wie folgt vor:

1. Je nach Spielsituation (*wieviele Münzen liegen noch auf dem Tisch?*) wählt er einen Zug aus dem zugehörigen mittleren Feld aus.
2. Die entsprechende Marke verschiebt er aus dem mittleren Feld in das rechte Feld. Alle anderen Marken in der letzten Spalte werden in die jeweils zugehörige mittlere Spalte zurückgelegt, da es in diesem Moment keinen Hinweis gibt, dass der dort gemerkte Zug ungeeignet war.
3. Anschließend führt er den Zug aus, nimmt also entsprechend der gewählten Marke Münzen vom Tisch.
4. Hat er jetzt verloren, entfernt er die gemerkte Marke aus dem Spiel.

Liegt aktuell in dem mittleren Feld keine Marke, macht er einen beliebigen Zug, indem er eine oder zwei Münzen entfernt. Die zuvor gemerkte Marke wird aus dem Spiel entfernt.

Die Lernstrategie am Beispiel erläutert

Die ersten beiden Runden (s.o.) verlaufen dann wie folgt:

- Es liegen 5 Münzen auf dem Tisch;



- Alice beginnt und nimmt eine Münze.



- Jetzt ist Bob an der Reihe. Er beschließt, zwei Münzen zu nehmen und vermerkt es dann in seiner Tabelle:

Zustand	geeignete bzw. sinnvolle Züge	Gemerker Zug
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm eine Münze</div>	<div>Nimm zwei Münzen</div>
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm eine Münze</div>	

- Es liegen dann noch zwei Münzen auf dem Tisch und Alice ist am Zug.



- Sie nimmt eine Münze, so dass Bob jetzt nur noch eine Münze vorfindet.





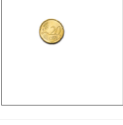


- Bob merkt noch nicht sofort, dass er verloren hat. denn erkannt ja noch einen Zug machen! Der zuletzt gemerkte Zug kann also zurückgelegt werden, da er noch nicht als *ungeeignet* erkannt wurde. Stattdessen wird er in jetzt dieser Spielsituation eine Münze nehmen und sich das merken:





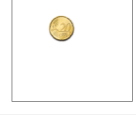
Zustand	geeignete bzw. sinnvolle Züge	Gemerker Zug
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
		<div>Nimm eine Münze</div>

•














- Jetzt merkt Bob, dass er die letzte Münze genommen hat: Er hat leider verloren! Die zuletzt gemerkte Marke entfernt er aus dem Spiel, da dieser Zug zum Verlust geführt hat:

Zustand	geeignete bzw. sinnvolle Züge	Gemerakter Zug
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
		













- Die nächste Runde kann beginnen. Bob hat den ersten Zug und nimmt zwei Münzen:

Zustand	geeignete bzw. sinnvolle Züge	Gemerakter Zug
	<div>Nimm eine Münze</div>	<div>Nimm zwei Münzen</div>
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
	<div>Nimm zwei Münzen</div> <div>Nimm eine Münze</div>	
		

- Alice findet noch 3 Münzen vor und nimmt zwei Münzen:

Zustand	geeignete bzw. sinnvolle Züge	Gemerker Zug
		
	 	
	 	
	 	
		

- Jetzt ist also Bob am Zug, findet jedoch keine passende Marke in dem mittleren Feld. Das bedeutet, dass der zuletzt gemerkte Zug zu einer Spielsituation führt, die keine Aussicht auf Erfolg für ihn hat. Er entfernt also diese Marke:

Zustand	geeignete bzw. sinnvolle Züge	Gemerker Zug
		
	 	
	 	
	 	
		

- Jetzt muss er den einzigen Zug machen, der möglich ist, also die letzte Münze nehmen, so dass er (mal wieder) verliert.

Was hat Bob bisher gelernt?

1. Er verliert immer dann, wenn nur noch eine Münze auf dem Tisch liegt.
2. Es nicht klug, wenn Bob zwei Münzen nimmt, wenn er fünf Münzen vorfindet.

Wie geht es weiter?

Bob sollte jetzt noch weitere Runden spielen. Immer, wenn Bob verliert, wird er sozusagen bestraft, indem die zuletzt gemerkte Marke entfernt wird. Ansonsten wird er dadurch belohnt, dass die zuletzt gemerkte Marke zurückgelegt wird, da der zugehörige Zug (noch) nicht als schlecht erkannt wurde.

Irgendwann wird die Tabelle sich nicht mehr ändern; Bob hat dann gelernt, in jeder Situation den besten Zug zu machen.

Aber Vorsicht! Es gibt Spielsituationen, in denen kein “bester Zug” in der mittleren Spalte zu finden ist. In dieser Situation kann er nicht mehr aus eigener Kraft gewinnen. Er muss dann einen Zufallszug machen und hoffen, dass Alice noch einen Fehler macht.

Beobachtung nach ausreichend langer Lernphase

- Es gibt keine Situation, in der in der mittleren Spalte zwei Marken liegen.
- In einigen Situationen gibt es genau einen möglichen Zug, der garantiert zum Sieg führt.
- Falls es in der mittleren Spalte keine Marke gibt, kann Bob nur noch gewinnen, wenn Alice anschließend einen Fehler macht.

Letzte Anmerkung

Natürlich kann man auch mit mehr als fünf Münzen starten, jedoch dauert dann die Lernphase entsprechend länger.

Bemerkung zu der Python-Umsetzung

Bob könnte ohne Benutzer-Interaktion die Züge machen, also einen Zufallszug, falls das Aktionsfeld leer ist, oder eine zufällige Aktion aus der Liste der Aktionen des Aktionsfeldes wählen.

Auch die Rolle von Alice könnte ohne Interaktion implementiert werden. Sie müsste dann ebenfalls zufällige Züge machen, jedoch ohne dabei zu lernen. Dann ist es wahrscheinlicher, dass alle möglichen Situationen durchgespielt werden.

Jedoch - aus methodisch-didaktischer Sicht - soll der Benutzer die Rolle von Bob annehmen, um zu sehen, was in den Aktionsfeldern geschieht.