Grundlagen der Künstlichen Intelligenz

Prof. Dr. J. Boedecker, Prof. Dr. W. Burgard, Prof. Dr. F. Hutter, Prof. Dr. B. Nebel, Dr. rer. nat. M. Tangermann
M. Krawez, T. Schulte
Sommersemester 2019

Universität Freiburg
Institut für Informatik

Übungsblatt 6 — Lösungen

Aufgabe 6.1 (Value-Iteration-Algorithmus)
Betrachten Sie die folgende Gitterwelt. Die u-Werte stehen für den Nutzen eines Zustandes, nachdem die *Value Iteration* konvergiert ist, r für die Belohnung, die ein Zustand erbringt. Nehmen Sie einen Discountfaktor $\gamma = 1$ an. Der Agent kann vier mögliche Aktionen ausführen: **Nord**, **Süd**, **Ost** und **West**. Mit Wahrscheinlichkeit 0,7 erreicht der Agent den Zustand, den er erreichen will, mit Wahrscheinlichkeit 0,2 bewegt er sich nach rechts und mit Wahrscheinlichkeit 0,1 nach links von der beabsichtigten Richtung.

$u = 8$	$u = 15$	$u = 12$
$u = 2$	$r = 2$	$u = 10$
$u = 7$	$u = 16$	$u = 11$

Welches ist die beste Aktion, die ein Agent ausführen kann, der sich im zentralen Zustand der Gitterwelt aufhält? Erklären Sie Ihre Antwort. Welchen Nutzen hat der zentrale Zustand damit?

Lösung:
Sei s der zentrale Zustand. Wir müssen für jede Aktion a den erwarteten Nutzen der Anwendung von a in s berechnen und dann die Aktion mit dem höchsten erwarteten Nutzen auswählen. Der erwartete Nutzen ist

$$\sum_{s'} T(s, a, s')U(s').$$

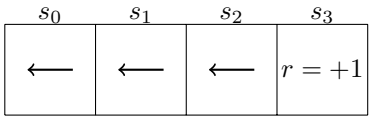
Konkret bedeutet das hier:

$$\begin{aligned}
\sum_{s'} T(s, \textbf{Ost}, s')U(s') &= 0{,}7 \cdot 10 + 0{,}2 \cdot 16 + 0{,}1 \cdot 15 = 7 + 3{,}2 + 1{,}5 = 11{,}7 \\
\sum_{s'} T(s, \textbf{Süd}, s')U(s') &= 0{,}7 \cdot 16 + 0{,}2 \cdot 2 + 0{,}1 \cdot 10 = 11{,}2 + 0{,}4 + 1 = 12{,}6 \\
\sum_{s'} T(s, \textbf{West}, s')U(s') &= 0{,}7 \cdot 2 + 0{,}2 \cdot 15 + 0{,}1 \cdot 16 = 1{,}4 + 3 + 1{,}6 = 6 \\
\sum_{s'} T(s, \textbf{Nord}, s')U(s') &= 0{,}7 \cdot 15 + 0{,}2 \cdot 10 + 0{,}1 \cdot 2 = 10{,}5 + 2 + 0{,}2 = 12{,}7
\end{aligned}$$

Also ist die beste Aktion $\arg\max_a \sum_{s'} T(s,a,s')U(s')$ ist also **Nord**. Der zentrale Zustand hat damit Nutzen $2 + 12{,}7 = 14{,}7$.

Aufgabe 6.2 (Policy-Iteration-Algorithmus)

Sei nun der Discount $\gamma = 0{,}5$ und die einzigen Aktionen seien **Ost** und **West**. Mit Wahrscheinlichkeit 0,9 erreicht der Agent den Zustand, den er erreichen will (bzw. bleibt stehen, falls die Aktion ihn über den Rand des Gitter hinausführen würde), und mit Wahrscheinlichkeit 0,1 bewegt er sich in die entgegengesetzte Richtung. Die Belohnung in den drei westlichen Zuständen ist jeweils $-0{,}05$.



Führen Sie einen Schritt der *Policy Iteration* durch, wobei die initiale Policy π_0 durch die Pfeile in den Zuständen gegeben ist. Geben Sie das lineare Gleichungssystem für die erste *Policy Evaluation* und eine Lösung des Gleichungssystems sowie die erste verbesserte Policy π_1 an.

Lösung:

Lineares Gleichungssystem:

$$\begin{aligned} U(s_0) &= -0.05 + 0.5(0.9U(s_0) + 0.1U(s_1)) \\ U(s_1) &= -0.05 + 0.5(0.9U(s_0) + 0.1U(s_2)) \\ U(s_2) &= -0.05 + 0.5(0.9U(s_1) + 0.1U(s_3)) \end{aligned}$$

bzw., ausmultipliziert:

$$\begin{aligned} U(s_0) &= -1/11 + 1/11U(s_1)) \\ U(s_1) &= -1/20 + 9/20U(s_0) + 1/20U(s_2) \\ U(s_2) &= 9/20U(s_1) \end{aligned}$$

Löst man dieses Gleichungssystem, so erhält man:

$$\begin{aligned} U(s_0) &= -\frac{411}{4121} \approx -0,0997 \\ U(s_1) &= -\frac{400}{4121} \approx -0,0971 \\ U(s_2) &= -\frac{180}{4121} \approx -0,0437 \end{aligned}$$

Die erste verbesserte Policy ist also

$$\begin{aligned} \pi_1(s_0) &= \textbf{Ost} \\ \pi_1(s_1) &= \textbf{Ost} \\ \pi_1(s_2) &= \textbf{Ost} \end{aligned}$$

Aufgabe 6.3 (Entscheidungsbaum-Lernen)

Zwei Kandidaten O und M, die mit ihren Programmen unterschiedliche Teile der Bevölkerung ansprechen, bewerben sich um ein politisches Amt. Die folgende Tabelle zeigt die Präferenzen von sieben Wählern mit unterschiedlichem Alter, Einkommen und Bildungshintergrund.

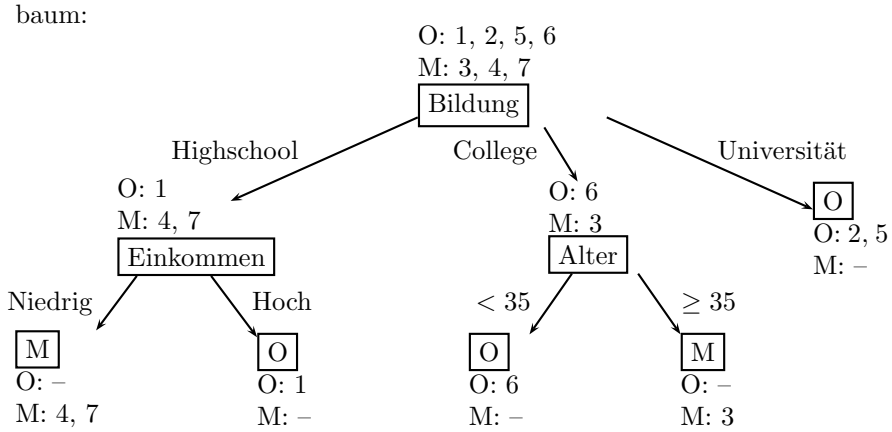
Nr.	Alter	Einkommen	Bildung	Kandidat
1	≥ 35	Hoch	Highschool	O
2	< 35	Niedrig	Universität	O
3	≥ 35	Hoch	College	M
4	≥ 35	Niedrig	Highschool	M
5	≥ 35	Hoch	Universität	O
6	< 35	Hoch	College	O
7	< 35	Niedrig	Highschool	M

(a) Berechnen Sie mit Hilfe des Lernalgorithmus aus der Vorlesung einen möglichst kleinen Entscheidungsbaum, der alle gegebenen Beispiele anhand der Attribute *Alter*, *Einkommen* und *Bildung* korrekt danach klassifiziert, welcher Kandidat bevorzugt wird. Geben Sie für den Wurzelknoten die *information gains* aller Kandidaten-Attribute an.

(b) Leiten Sie aus dem Entscheidungsbaum eine logische Formel ab, die genau dann erfüllt ist, wenn Kandidat O bevorzugt wird.

Lösung:

(a) Der Algorithmus aus der Vorlesung findet den folgenden Entscheidungsbaum:



Da am Wurzelknoten vier zu drei Einträge stehen, also $\frac{p}{p+n} = \frac{4}{7}$, $\frac{n}{p+n} = \frac{3}{7}$, ist der Informationsgehalt am Wurzelknoten $I(\frac{p}{p+n}, \frac{n}{p+n}) = I(\frac{4}{7}, \frac{3}{7}) \approx 0{,}985$. Die möglichen Attribute für die erste Verzweigung sind *Alter*, *Einkommen* und *Bildung*.
Verzweigung nach *Alter*:

$$R(Alter) = \underbrace{\frac{3}{7} \cdot I(\frac{2}{3}, \frac{1}{3})}_{Jung(<35)} + \underbrace{\frac{4}{7} \cdot I(\frac{2}{4}, \frac{2}{4})}_{Alt(\geq 35)}$$

$$\approx 0{,}394 + 0{,}571 = 0{,}965$$

$$Gain(Alter) \approx 0{,}985 - 0{,}965 = 0{,}020$$

Verzweigung nach *Einkommen*:

$$\begin{aligned} R(Einkommen) &= \underbrace{\frac{4}{7} \cdot I(\frac{3}{4}, \frac{1}{4})}_{Hoch} + \underbrace{\frac{3}{7} \cdot I(\frac{1}{3}, \frac{2}{3})}_{Niedrig} \\ &\approx 0{,}464 + 0{,}394 = 0{,}857 \\ Gain(Einkommen) &\approx 0{,}985 - 0{,}857 = 0{,}128 \end{aligned}$$

Verzweigung nach *Bildung*:

$$\begin{aligned} R(Bildung) &= \underbrace{\frac{3}{7} \cdot I(\frac{1}{3}, \frac{2}{3})}_{Highschool} + \underbrace{\frac{2}{7} \cdot I(\frac{1}{2}, \frac{1}{2})}_{College} + \underbrace{\frac{2}{7} \cdot I(\frac{2}{2}, \frac{0}{2})}_{Universität} \\ &\approx 0{,}394 + 0{,}286 + 0 = 0{,}679 \\ Gain(Bildung) &\approx 0{,}985 - 0{,}679 = 0{,}306 \end{aligned}$$

Wegen $Gain(Bildung) > Gain(Einkommen) > Gain(Alter)$ liefert also ein Test bzgl. *Bildung* den größten Informationsgewinn, man sollte also am Wurzelknoten zuerst bzgl. *Bildung* testen.

(b)

$$\begin{aligned} O \quad \equiv \quad & Bildung = Universität \vee \\ & (Bildung = College \wedge Alter < 35) \vee \\ & (Bildung = Highschool \wedge Einkommen = Hoch) \end{aligned}$$