

MMKI-Praktikum SS06: Aufgabe 4D

Die Aufgabe 4D schließt nahtlos an die Aufgaben aus 3D an.

Ziel dieses Aufgabenblattes ist die weiterführende Beschäftigung mit dem von ihnen gewählten Lernverfahren.

Dazu sollen die zwei Teilaufgaben aus 3D verknüpft und durch einen gemeinsamen Controller gelöst werden, der wieder durch die Anwendung eines Lernverfahrens gefunden werden soll.

Aufgabe:

- Bringen Sie den Simloid in die bekannte Ausgangslage (alle Gelenke mit dem PID-Controller fixiert, beide Arme zeigen senkrecht nach unten), sobald alle Gelenke ihre Zielstellung erreicht haben, schalten Sie die Reibung für das rechte Armgelenk (und nur für dieses) mit dem friction-Kommando aus.
- Erzeugen Sie einen Controller durch ein maschinelles Lernverfahren (vorzugsweise sollten Sie das von Ihnen in der vergangenen Aufgabe verwendete benutzen), der den rechten Arm des Simloids deterministisch in die Senkrechte (nach oben!) bewegt. und dort für mindestens 5s balanciert (maximale Auslenkung um den oberen Totpunkt: 5°).
- Es gelten die selben Einschränkungen bzgl. des maximalen torque, wie im letzten Aufgabenblatt, d.h. nach der Initialisierung sollen nur noch (torque $2 <x>$) Kommandos abgesendet werden, der Betrag des Torques soll 30 nicht überschreiten.
- Die Aufgabe ist **nicht** gelöst, wenn Sie die zwei Controller der letzten Aufgabe nacheinander ausführen! Ziel ist genau **ein** Controller, der die gesamte Aufgabe löst.
- Sie können wieder aus verschiedenen Optimierungskriterien auswählen:
 - Minimieren Sie die Gesamtzeit bis zum Erreichen der stabilisierten Endposition.
 - Minimieren Sie den maximal verwendeten torque-Betrag.
(Selbstverständlich sollte das Erreichen des eigentlichen Ziels noch in vertretbarer Zeit geschehen)
 - Minimieren Sie die durch den Controller benutzte Energie (Zeitintegral der torque-Beträge)

Abgabeformat:

Neben der Programmierung / Konfiguration des Lernverfahrens und dessen Anwendung zur Lösung der Aufgabe sind folgende Punkte ebenfalls Teil der Aufgabe: (bitte kurz und wesentlich behandeln)

- Beschreiben Sie kurz das von ihnen gewählte Lernexperiment und den Verlauf der Lernphase. Begründen Sie Ihre Wahl.
Im Fall von Reinforcement Learning-Verfahren sind dabei auch die Parameter wie Lernrate, Trace-Value, Lernpolitik, ... relevant.
Sofern Sie sich für ein besonders neues oder trickreiches Verfahren entschieden haben, erläutern Sie es kurz.
- Beschreiben und begründen Sie den von Ihnen modellierten Zustandsraum.
Welche Zustandsvariablen (Dimensionen) gibt es, wie werden diese berechnet, wie ist der Zustandsraum aufgebaut?
Verwenden Sie einen diskreten Zustandsraum, ist natürlich auch die Form der Diskretisierung wichtig.
- Beschreiben und begründen Sie den von Ihnen modellierten Aktionsraum.
Welche Aktionstypen oder konkrete Aktionen gibt es, wie werden diese ggf. in die nativen Simloid-Kommandos umgerechnet?
Verwenden Sie einen diskreten Aktionsraum, ist wiederum die Form der Diskretisierung relevant.
- Beschreiben und begründen Sie das von Ihnen als Ziel des Lernprozesses gewählte Modell.
Dies kann zum Beispiel ein Entscheidungsbaum, ein Neuronales Netz, ein Fuzzy-Controller oder im Fall von Reinforcement Learning eine beliebige Representation der Wertfunktion sein (z.B. LookUp-Tabelle oder TileCodings).
- Sollten Sie, was durchaus erwünscht ist, verschiedene Verfahren, Modellierungen und Experimente probiert haben, um zu einer Lösung zu gelangen, beschreiben Sie kurz ihre Herangehensweise und die Auswirkungen ihrer Änderungen.

Um den Lernerfolg zu dokumentieren, sind mind. 2 graphische Plots abzugeben:

- Für die Lernphase ist ein Diagramm zu erstellen, welches ein geeignetes Maß für den Lernerfolg gegen die Zeit oder die Anzahl der Takte oder Episoden aufträgt.
Die Zeitskala soll bei Null starten und bis zum Erreichen eines hinreichend stabilen Verhaltens gehen.
- Der gelernte Controller soll fünf mal ausgehend vom Initialzustand aufgerufen werden, um die korrekte Lösung zu dokumentieren.
Für diese Episoden soll jeweils der Gelenkwinkel des rechten Schultergelenks in ein gemeinsames Diagramm gezeichnet werden. (x-Achse ist wieder die Zeit in s oder Takten)

Abzugeben sind ferner die Quellcodes der Programme, sowie zwei Startskripte:

1. Lernphase: Start des Simloid und des Controllers, der die Struktur lernt.
2. Testphase: Start des Simloid und des Controllers mit der gelernten Struktur.

Hinweise:

Sie werden feststellen, dass diese nur leicht veränderte Lernaufgabe schon etwas schwieriger zu lösen ist, beginnen Sie daher rechtzeitig mit ihren Experimenten.

Das von Ihnen gewählte Experiment, die Größe und die genaue Modellierung des Zustands- und des Aktionsraumes haben signifikanten Einfluß auf den Lernerfolg und die Lerngeschwindigkeit (in beide Richtungen)!

Wenn ihr Lernsystem zwar etwas lernt (konvergiert), aber nicht das, was es soll, ist in aller Regel das von Ihnen gewählte feedback-Signal ungünstig gewählt, ihr Zustands/- Aktionsraum unzureichend oder ihr Lernexperiment der Aufgabe nicht angemessen. Experimentieren Sie!

weitere Hinweise:

Wie bereits mehrfach erwähnt, ist das Winkelmodell der Dynamixels unvollständig und unstetig.

Insbesondere gibt es einen jeweils 30° breiten Bereich, der nur durch die Werte 0 und 1023 beschrieben ist.

Sie können es dem Lernsystem leichter machen, indem Sie ihm Informationen bereitstellen, um innerhalb dieses Bereiches trotzdem zu differenzieren, z.B. in dem Sie Beschleunigungsdaten in den Zustandsraum aufnehmen.

Auch müssen Sie die zur Verfügung stehenden Aktionen nicht auf die beiden (torque 2 30), (torque 2 -30) beschränken, experimentieren Sie mit mehr Aktionen oder komplexeren, so genannten Metaaktionen.

Falls Sie Verfahren des Reinforcement Learning verwenden, könnten folgende Tipps bei der Lösung helfen.

- Mitunter ist es hilfreich während des Lernprozesses die Reward-Funktion und/oder das Experiment **leicht!** zu verändern, wenn man bereits gelerntes Verhalten zum Teil erhalten möchte.
- Mit Layered Learning bezeichnet man die Dekomposition von komplexen Lernaufgaben in kleinere separat lernbare Aufgaben. Sie können etwas ähnliches realisieren, indem Sie ihrem Controller z.B. zuerst „beibringen“, den Arm zu balancieren und erst nachdem dies funktioniert, lernen Sie das Aufschaukeln (oder umgekehrt).
Variieren Sie dazu das Lernexperiment entsprechend, ohne bereits gelerntes Verhalten zu „zerstören“.