
Aufgabenstellung

Am Fachgebiet wurde vor einiger Zeit ein aktuiertes Doppelpendel als Demonstrator für das Matlab II Praktikum angefertigt. Es stellte sich bei vergangenen Arbeiten heraus, dass die *Stabilisierung der oberen Ruhelage* aufgrund einer *Lose* im Antriebszweig nicht trivial ist. Die vergangenen Arbeiten beschäftigten sich mit *traditionellen* Regelungen wie beispielsweise einem *Gain Scheduling Regler*. Zudem wurde auch der Versuch unternommen diese *Lose* zu kompensieren und das System wurde auf Grenzyklen untersucht. Trotzdem blieb immer eine ausgeprägte Dauerschwingung im geregelten System zurück.

Ziel dieser Masterarbeit ist es, mittels *Reinforcement Learning* (RL) einen Regler für das oben beschriebene System zu lernen. Hierfür soll zunächst nach geeigneten Methoden zur Lösung dieser Aufgabe recherchiert werden. Diese sollen dann in einer Simulation getestet und bewertet werden. Zudem sollen auch traditionelle Regelungsmethoden als Baseline implementiert und getestet werden. Zu Beginn kann ein stark vereinfachtes Modell betrachtet werden, welches im Laufe der Arbeit weiter ergänzt wird. Als Basis bietet sich hier das Model des Matlab II Praktikums an, wobei der Zustand als messbar angenommen werden kann. In weiteren Schritten sollen dann die folgenden Ergänzungen vorgenommen werden:

- Nur die Winkelpositionen sind messbar
- Anstelle eines Drehmoments wird eine Motorspannung vorgegeben
- Die *Lose* und ggf. auch Coulombsche-Reibung sollen in das Modell mit aufgenommen werden
- Rechenzeit und Trainingsdauer müssen berücksichtigt werden.

Werden die obigen Punkte erfolgreich bearbeitet, dann sollen die entworfenen Regelungen am realen System getestet werden. Alle Ergebnisse (einschließlich des erzeugtem Programmcodes) sind sorgfältig zu dokumentieren. Des Weiteren sollen die Ergebnisse falls möglich geeignet visualisiert werden.

Zusätzlich zur obigen Aufgabenstellung können folgende Leitfragen näher betrachtet werden

- Wie kann RL für kontinuierliche *action spaces* genutzt werden? Welche Methoden gibt es?
- Kann das Wissen über ein approximatives Modell sinnvoll genutzt werden? Falls ja, wie?
- Kann ein Baseline Regler die Trainingsdauer verkürzen?
- Können Stabilitäts- bzw. Performance-Garantien gegeben werden?
- Wie sehen mögliche *Rewards* aus? Führen unterschiedliche Rewards auf dieselbe optimale Lösung?
- Welche Möglichkeiten gibt es die *Observations* darzustellen? Kann hier ein Beobachter das Ergebnis verbessern?
- Wie wirken sich *Time delays* und *Rechenzeiten* auf das System aus?

Diese Fragen müssen nicht vollständig beantwortet werden und dem Studierenden steht es offen sich einige Fragestellungen auszuwählen oder auch eigene zu formulieren.

Beginn: _____
Ende: _____
Seminar: _____