Praktikum Autonome Roboter

—

Projekt IV
Erwerb von Verhaltenskompetenzen durch Reinforcement–Lernen

Priv.-Doz. Dr. R. Der
Institut für Informatik
Universität Leipzig

30. September 1997

1 Aufgabenstellung

Das Projekt ähnelt dem Projekt III stellt aber geringere Ansprüche, da keine visuellen Daten von der Kamera verarbeitet werden müssen. Aufgabe des AR ist wieder, eine bestimmte Verhaltensleistung zu erlernen. Dabei können die gleichen Lernverfahren wie im Projekt III eingesetzt werden. Es soll deshalb hier nur die Aufgabenstellung und die im Vergleich zu Projekt III andere Diskretisierung des Zustands– und Aktionsraumes dargestellt werden.

Aufgabe: Der autonome Roboter (AR) soll eine der folgenden Verhaltenskompetenzen durch **Reinforcement–Lernen** erwerben:

1. Der AR soll lernen, im freien Raum eine runde Scheibe vor sich her zu schieben[1], so daß er im wesentlichen eine gerade Strecke zurücklegt. Das entspricht in gewisser Weise der Lernaufgabe für den Stabbalancer, die in der dynamischen Optimierung eine Benchmarkaufgabe darstellt, aber auch in der neuronalen Netzwerkliteratur oft behandelt wird, vgl. [3].

2. Der AR soll lernen, sich in einer komplizierten Umgebungsgeometrie (Labyrinth) sicher zu bewgen, d. h. er soll möglichst große Strecken zurücklegen können, ohne anzuecken oder festzufahren. Als Anfangsvariante kann man auch ein einfaches Wandfolgeverhalten anstreben.

[1] Wenn die Aufgabe mit dem Simulator angegangen werden soll, so muß zunächst die Dynamik der Scheibe integriert werden.

Im Praktikum ist eines der beiden im Projekt III geschilderten Lernverfahren einzusetzen. Alternativ kann auch ein konservativer, d. h. regelbasierter Ansatz versucht und die Ergebnisse miteinander verglichen werden.

Weiterführende Literatur: Vorlesungsskript "Robotik" [1].

2 Diskretisierung von Zustands– und Aktionsraum

Die einzusetzenden Lernverfahren setzen eine geeignet diskretisierte Darstellung des Zustands–Aktionsraumes voraus:

2.1 Der Zustandsraum

Im Unterschied zu Projekt III liegen hier sensorische Inputs nur von den Infrarotsensoren vor, bei der Aufgabenstellung 1 ist die einzige relevante Information für das Verhalten des AR die Stellung der Scheibe relativ zum Roboter, die Richtung, in die die Scheibe geschoben werden soll ist nicht vorgegeben. Wichtig ist nur, daß die Bewegung möglichst geradlinig erfolgt. (Man kann sich natürlich auch eine andere geometrische Figur als Zielbewegung vorgeben).

Zur Diskretisierung wird das Grundintervall des jeweiligen Sensors in mehrere Teilintervalle unterteilt. Diese Aufteilung ist für die Sensoren je nach Lage unterschiedlich zu wählen. Für die seitlichen (senkrecht zur Nase des Khepera zeigenden) Sensoren reicht sicher die Unterteilung in zwei Intervalle, für die zwei frontalen empfiehlt sich eine feinere Diskretisierung. Diese sollte zudem nicht äquidistant sein, sondern den Bereich höherer Sensorwerte feiner unterteilen.

Für den Anfang sollte der Zustandsraum aber nicht groß sein, Richtwert ist $6-10$ Zustände. **Dabei ist die Symmetrie des Problems geeignet auszunutzen**.

2.2 Der Aktionsraum

Um die Komplexität der Lernaufgabe klein zu halten, muß die Zahl der möglichen Aktionen klein gehalten werden. Für die Aufgabe 1 kommt man für den Anfang sicher mit $3-5$ Aktionen aus, diese können sein:

- Geradeausfahren
- Fahren auf einem Kreisbogen mit jeweils vorgegebenem Radius, dabei wieder feinere Unterteilung für die Aktionen, die zu kleinen Winkeländerungen (großer Radius) führen.

Für den Fall der Aufgabe 2 sollte man unbedingt auch eine Aktion haben, die die letzten ein bis drei Bewegungen in umgekehrter Reihenfolge zu durchlaufen gestattet (Herausfinden aus einer Sackgasse). Dazu sind die Steuerungsanweisungen zu speichern. Man beachte, daß ein Update der Q– bzw. der Wertfunktion nur dann erfolgt, wenn sich der Zustand im **diskretisierten** Zustandsraum ändert.

3 Lernverfahren

Die Lernverfahren sind im Projekt III beschrieben, weitere Ausführungen im Skript zur Vorlesung Robotik [2].

Aufgabenspezifisch ist nur die Zuweisung des Reinforcements, für das man sich ein geeignetes Scenario überlegen muß%. Vor allem muß das möglichst automatisiert erfolgen. Im Fall der Aufgabe 1 kann man sich vorstellen, daß der AR immer dann Reinforcement erhält, wenn die Scheibe einigermaßen zentral vor ihm liegt, negatives Reinforcement, wenn er die Scheibe verloren hat oder diese vor die seitlichen Sensoren zu liegen kommt. Das Reinforcement kann folglich als eine Funktion der Sensorwerte automatisch bestimmt werden.

Analoge Überlegungen gelten für Aufgabe 2.

4 Weiterführende Arbeiten

Methoden zur Effektivierung der Lernverfahren wurden in Projekt III diskutiert. Eine Leistungsverbesserung des angelernten AR ist vor allem durch Überwindung der Diskretisierungseffekte möglich. Das kann durch eine Interpolation zwischen den diskreten Zuständen und Aktionen erreicht werden. Das ist aber eine nichttriviale Angelegenheit!

Literatur

[1] R. Der. Vorlesung Robotik. Universität Leipzig Institut für Informatik. http://www.informatik.uni-leipzig.de/ der/Vorlesungen/skrpt-rob.html.

[2] R. Der. Vorlesung Robotik. UniversitätLeipzig Institut für Informatik Kap. 2.3: Die Planungskomponente. http://www.informatik.uni-leipzig.de/ der/Vorlesungen/rob-plan.ps.Z.

[3] H. J. Ritter, T. M. Martinetz, and K. J. Schulten. *Neuronale Netze: Eine Einführung in die Neuroinformatik selbstorganisierender Abbildungen.* Addison-Wesley, Munich, Germany, 1990.