Synthese von Steuerprogrammen durch Klassifizierungslernen am Beispiel der Stabilisierungssteuerung von Nachrichtensatelliten

Wolfgang Müller, FhG,IITB,EPO,Kurstr 33, D-10117 Berlin

Fritz Wysotzki, TU Berlin, Fachbereich Informatik, Franklinstr. 28/29, D-10587 Berlin

Christian Brühe, TU Berlin, Fachbereich Informatik, Franklinstr. 28/29, D-10587 Berlin

email: wmueller@epo.iitb.fhg.de

Einleitung

Verfahren des Maschinellen Lernens haben heute eine Reife erreicht, die zu ersten erfolgreichen industriellen Anwendungen geführt hat. In der Prozeßdiagnose und -steuerung ermöglichen Lernverfahren die Klassifikation und Bewertung von Betriebszuständen, d.h. eine Grobmodellierung eines Prozesses, wenn dieser nicht oder nur teilweise mathematisch beschreibbar ist. Außerdem gestatten Lernverfahren die automatische Generierung von Klassifizierungsprozeduren, die deterministisch abgearbeitet werden und daher für die Belange der Echtzeitdiagnose und -steuerung u.U. zeiteffektiver als Inferenzmechanismen auf logischer bzw. Produktionsregelbasis sind, da letztere immer mit zeitaufwendigen Suchprozessen verbunden sind.

In den vergangenen Jahren sind in der KI und Neuroinformatik eine Reihe von Lernverfahren zur Generierung von lokal optimalen Steuertrajektorien in Echtzeit untersucht worden (z. B. BOXES [4], CART [3], ASE/ACE [1]). Ziel dieser heute sehr aktuellen Grundlagenforschung ist die Modellierung von Fertigkeiten. Hauptproblem ist hier das Zurückspielen der Erfolgs- oder Mißerfolgsrückmeldung, die erst nach einer Sequenz von Steuermaßnahmen erfolgt, auf die einzelnen, lokal generierten Steueroperationen (Credit-Assignment-Problem [8],[9]). Das meist prototypisch untersuchte (Spiel-) Beispiel ist das Stab- und Wagenproblem, das hier nicht näher erläutert werden soll. Diese Verfahren sind in der Echtzeitphase relativ zeitaufwendig und falls mehr als zwei diskrete Steuermaßnahmen vorhanden sind, steigt die Komplexität des Problems sehr stark an.

Nachstehend wird ein Ansatz zur Generierung von Steueraktionsfolgen durch Kombination von Problemlösungstechniken mit Lernverfahren beschrieben, der echtzeitfähig ist und eine größere Anzahl diskreter Steuermaßnahmen erlaubt.

Lösungsansatz einer Steuerung von Prozessen auf der Basis von Lernverfahren

Am Beispiel der Stabilisierung einer Achse eines Nachrichtensatelliten (in Bezug auf die Stellung zur Erdachse) wurde die Anwendung des Klassifizierungslernens zur Ableitung von Steuermaßnahmen zur Störungskompensation untersucht. Der reale Prozeß der Steuerung eines Satelliten bezüglich seiner Achsenabweichung wurde durch einen bei MBB im Rahmen des ESPRIT- Projektes STATLOG [5] entwickelten Simulator ersetzt.

Ausgewählten Punkten, des durch Winkel φ und Winkelgeschwindigkeit $\dot{\varphi}$ definierten Zustandsraums der Achse wird zunächst mit Hilfe eines einfachen Problemlösers eine optimale Steueraktion zugeordnet, die automatisch aus einer (endlichen) Menge von ausführbaren Steuermaßnahmen ausgewählt wird. Optimal soll bedeuten, daß mit Hilfe eines empirischen Gütefunktionals (empirisches Wissen) eine Bewertung der Steueraktionen erfolgt. Ist das Gütefunktional z.B. der euklidische Abstand vom Nullpunkt (angestrebter Sollwert, d.h. Normallage des Systems), dann wird diejenige Steueraktion ausgewählt, die nach Anwendung auf den gegenwärtigen Zustand den Abstand am stärksten verringert. Es wurden verschiedene konvexe, empirische Gütefunktionale untersucht und angewendet, so daß im vorliegenden Anwendungsfall durch die lokale Optimalität auch die globale Optimalität der durch Verkettung der lokalen Aktionen erzeugten Steuertrajektorien garantiert wird (Greedy-Verhalten).

In einigen Fällen wurden Parameter der Funktionale, z.B. das Verhältnis von Winkel- zu Winkelgeschwindigkeitssteuerung, durch theoretische Modellüberlegungen optimiert, da sich die Drehung des Satelliten durch einfache physikalische Grobmodelle beschreiben läßt (Kopplung von empirischem mit Modellwissen).

Durch die Zuordnung einer optimalen Steueraktion zu ausgewählten Punkten des Zustandsraums entstehen homogene Gebiete, denen die gleiche Steuermaßnahme zugeordnet wird (für die Qualität der lernenden Steuerung ist es sinnvoll, eine Gleichverteilung der Punkte im relevanten Teil des Zustandsraumes zu erreichen). Damit ist ein Klassifizierungsproblem gegeben, das mit Methoden des überwachten Lernens gelöst werden kann. Man könnte zunächst alle Paare *Zustand/zugeordnete Steueraktion* als eine Menge (Tabelle) von elementaren Regeln auffassen und, um für einen gestörten Zustand die zugeordnete optimale Steuermaßnahme zu erhalten, eine Tabellensuche mit Identifikation des zum eingegebenen Zustand nächstliegenden gespeicherten Zustandes durchführen.

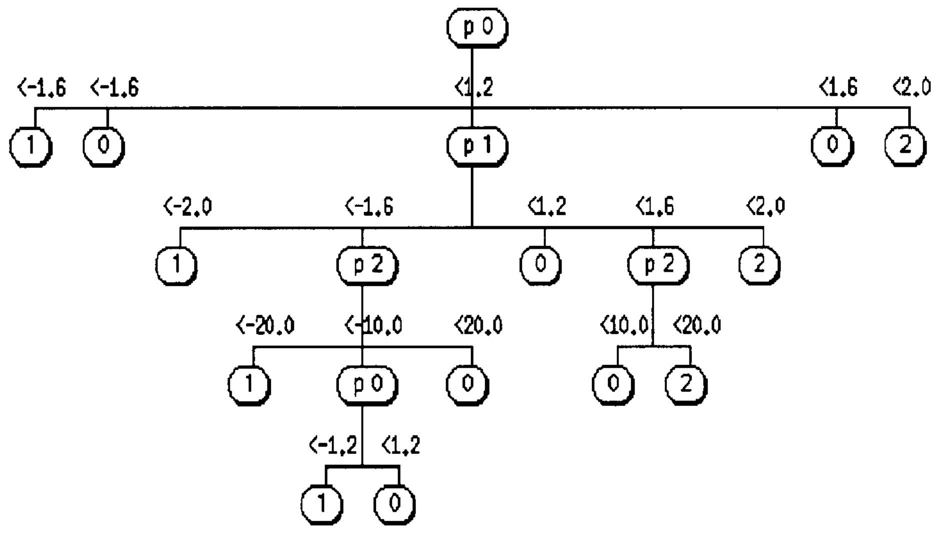

Abb. 1: Beispiel eines Entscheidungsbaums

Dieser Weg wurde experimentell untersucht, er führt zu unter Echtzeitbedingungen nicht mehr akzeptablen Rechenzeiten. Deshalb wurden die Lernverfahren CAL5 und DIPOL92 [5], [7], [10] eingesetzt, die anhand einer Lernmenge optimale Entscheidungsfunktionen für die Zuordnung eines eingegebenen gestörten Zustandes zu einer Klasse (Steuermaßnahme) entwickeln. Dadurch erfolgt gewissermaßen eine Verlagerung des Steuerproblems auf die höhere Hierarchieebene der Klasseneinteilung. Ein Beispiel des mittels CAL5 erzeugten Entscheidungsbaumes ist in Abb. 1 dargestellt.

Dieser so entwickelte Klassifikator kann als nichtlineares Steuerglied für den Prozeß verwendet werden. Mit dem z.B. in Abb. 1 gezeigten Entscheidungsbaum kann bei gegebenen Prozeßparametern durch sequentielle Abarbeitung der Tests die für diesen Prozeßzustand günstigste Steueraktion bestimmt werden. Durch rekursive Anwendung des Klassifikators kann man bei Vorgabe eines Zeittaktes die Steuertrajektorien im Zustandsraum des Prozesses als Sequenzen von lokaloptimalen Steuermaßnahmen (Endknoten des Entscheidungsbaumes) ermitteln. Die durch CAL5 und DIPOL92 entwickelten Klassifikatoren wurden als nichtlineare Regler in den Simulator eingebaut, für die Bewegung des Satelliten getestet und mit dem Verhalten eines im Simulator verfügbaren PD-Reglers verglichen.
In Abb. 2 ist das simulierte Steuerungsverhalten bei einer Störung von 0.05 Nm über 50 sec am Satelliten dargestellt. Die Störung setzt nach 3 sec ein, und der Übergang zum stabilen Zustand beginnt nach 53 sec, d.h. sofort nach dem Aussetzen der Störung. In Abb. 2 sind die Zeitverläufe der mit der Störung überlagerten Steuerimpulse [1](Tex), des Winkels (phi), der Winkelgeschwindigkeit (phi‘) und einer Kostenfunktion gezeigt. Die Kostenfunktion ist proportional dem Integral über den Zeitverlauf der Steuerfunktion und repräsentiert den Treibstoffverbrauch.

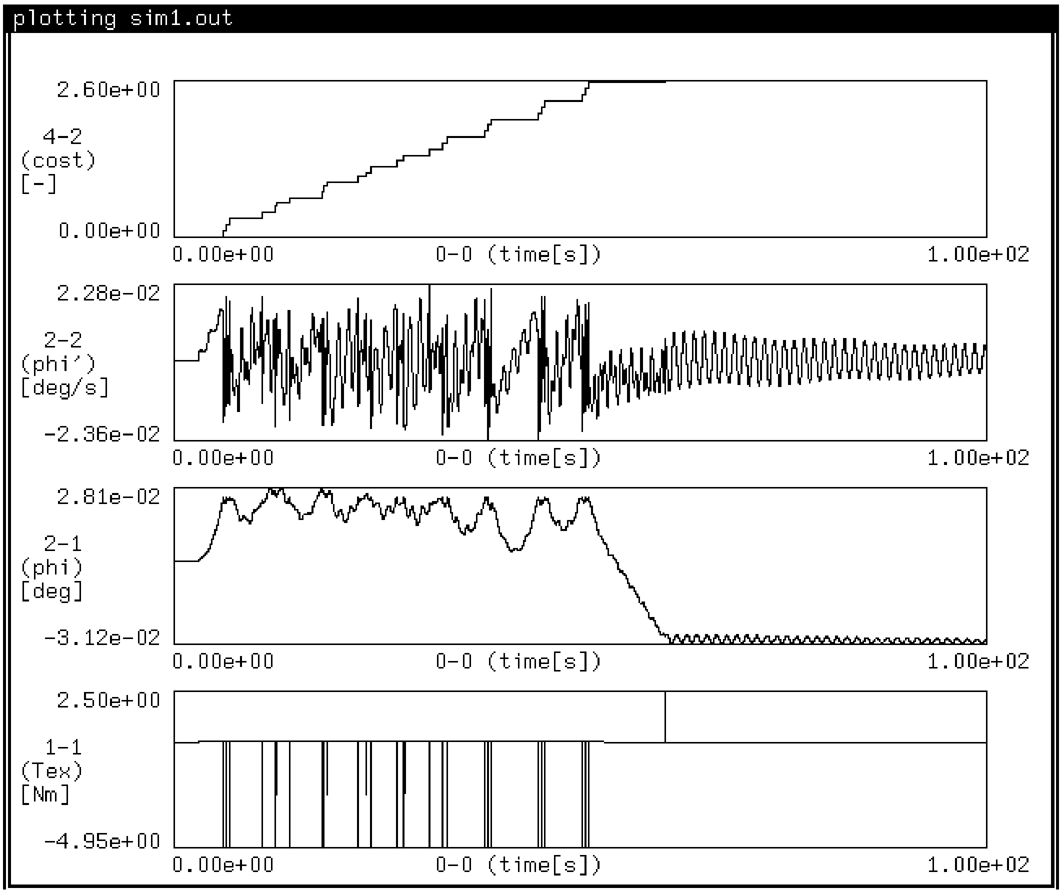

Abb. 2: Protokoll einer simulierten Satellitensteuerung mit Entscheidungsbäumen

Ein Problem war die Berücksichtigung von Koppelschwingungen, die durch die Solararrays

1. für diesen Versuch wurde mit 5 möglichen Steueraktionen (+5,+2.5,0,-2.5,-5 Nm) gearbeitet

des Satelliten hervorgerufen werden und die sich den zeitlichen Verläufen von Winkel und Winkelgeschwindigkeit überlagern. Ursprünglich wurden die Bewegungsparameter der Solararrays mit in die Zustandsdefinition aufgenommen. Es zeigte sich aber, daß sie sich durch eine Berücksichtigung einer gewissen „Historie“ des betrachteten Zustandes (zeitlich vorhergehende Zustände) im auf Winkel und Winkelgeschwindigkeit (der Drehung des Satelliten um die betrachtete Achse) reduzierten Zustandraumes kompensieren ließen. Zum anderen wurde festgestellt, daß die durch die Klassifikatoren erreichte Generalisierung der Klassen (Steuermaßnahmen) auch Schwankungen, hervorgerufen durch die Überlagerung von Koppelschwingungen auf die Drehbewegung des Satelliten, ausglich. Die Koppelschwingungen konnten daher als Rauschen aufgefaßt werden, das durch die Generalisierung zum größten Teil eliminiert wurde.

Schlußfolgerungen

Es wurde mit dem Simulator eine große Anzahl von Versuchen durchgeführt [6] [2], in denen z.B der Einfluß des Gütefunktionals und des Generalisierungsgrades der im Lernschritt erzeugten Klassifikatoren auf die Qualität der Steuerung untersucht wurden. Das Gütefunktional hat entscheidenden Einfluß auf den Treibstoffverbrauch und auf die Steuerungsqualität. Bei einem zu großem Generalisierungsgrad, der sich durch die Kompaktheit der Entscheidungsbäume widerspiegelt, verringert sich die Steuerungsqualität. Es kann sogar vorkommen, daß der zugelassene Winkel der Abweichung von der Satellitenachse bei zu kompakten Bäumen nicht mehr eingehalten werden kann.
Die Klassifikatorsteuerung war in vielen Fällen der PD-Steuerung überlegen in dem Sinne, daß näherungsweise stabile Zustände innerhalb des vorgegebenen Sollbereichs ohne Überschwingen erreicht bzw. weniger Steueraktionen (also geringerer Treibstoffverbrauch!) zur Erreichung des Sollbereichs benötigt wurden.

Die Funktionstüchtigkeit der oben beschriebenden lernenden Steuerung wurde mit Hilfe des Simulators der Satellitenbewegungen prinzipiell gezeigt. Bislang wurden Steuerungen untersucht, für die konvexe Funktionale benutzbar sind, so daß die erreichten lokalen optimalen Steuertrajektorien auch global optimal sind. Die Untersuchungen werden z.Zt. auf den allgemeineren Fall von Prozessen ausgedehnt, bei denen zur Erreichung des Sollbereichs im Zustandsraum bei der Steuerung eventuell vorhandene Instabilitätsgebiete umgangen werden müssen, also die empirische Bewertungsfunktion (Gütefunktional) nicht mehr konvex ist und mehrere Extrema haben kann.

Literatur

[1] Barto, A. G./Sutton, R. S./Anderson, C. W.:
Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems,
IEEE Transactions on Systems, Man and Cybernetics,
Vol. SMC-13, Number 5, pp 835 - 846, 1983.

[2] Brühe, C.:
Anwendung von KI-Methoden bei der Steuerung schwer modellierbarer kontinuierlicher Prozesse
Studienarbeit, Fachbereich Informatik, TU Berlin, 1993

[3] Connell, M. G./Utgoff, P. E.:
Learning to Control a Dynamic Physical System,
AAAI 87, Seattle, pp 456 - 460, 1987

[4] Michie, D./Chambers, R. A.:
BOXES as a Model of Pattern-formation,
Towards a Theoretical Biology, Vol. 1, Prolegomena,
C. H. Waddington (Ed.), Edinburgh: Edinburgh Univ. Press, pp 206 - 215, 1968.

[5] Michie, D./Spiegelhalter d.j./Taylor C.C.:
Machine Learning, Neural and Statistical Classification
Ellis Horwood, Hertfordshire, UK, 1994

[6] Müller, W./Schulmeister, B./Wisotzki, C./Wysotzki, F.:
BMFT-Verbundvorhaben WISCON, Teilvorhaben: Integration von Modellwissen, empirischem Wissen und Lernverfahren
Unveröffentlichter Forschungsbericht zu den Etappen MEL5 und MEL6, 1994

[7] Müller, W/Wysotzki, F.:
Automatic Construction of Decision Trees for Classification
in Annals of Operation Research 92,Moser, K./Schader M. (eds):
J.C. Baltzer AG Science Publishers, Wijdenes, Netherlands, 1994

[8] Sutton, R. S.:
Temporal Credit Assignment in Reinforcement Learning, Dissertation,
Departement of Computer and Information Science, University of Massachusetts, 1986.

[9] van Luenen, W. T. C.:
Real Time Reinforcement Learning Control of Dynamic Systems Applied to an Inverted Pendulum,
In:J. Stender, T. Addis (Eds.): Symbols versus Neurons,
IOS Press, Amsterdam, Washington, Tokyo, 1990, pp 232 - 247

[10] Unger, S/Wysotzki, F.:
Lernfähige Klassifizierungsysteme
Akademieverlag, Berlin, 1981