Operationale Beschreibungen des Lernens: Wenn Maschinen Begriffe erwerben

Katharina Morik
Univ. Dortmund, Computer Science Department, LS VIII
morik@ls8.informatik.uni-dortmund.de

Maschinenwelt – dabei denke ich nach einem kurzen Aufblitzen von Charlie Chaplin in "Moderne Zeiten" gleich an die Welten, die von Maschinen beschrieben werden. Maschinen, die – fast so wie ein Buch – etwas beschreiben, sind Rechenanlagen bzw. die Programme, die auf ihnen operieren. Turkle nannte sie "Wundermaschinen" [20], denn im Gegensatz zu den Beschreibungen von Sachverhalten und Ereignissen, die Bücher mithilfe von nicht mehr als 90 Zeichen in Form von Druckerschwärze auf Papier darstellen, *tun* die Beschreibungen, die Rechenanlagen mithilfe von nur zwei Zeichen in Form von elektrischer Spannung darstellen, gerade das, was sie beschreiben. Rechenanlagen führen eine Beschreibung aus, die Beschreibung ist operational. Beschreibt man formal, wie der Vorgang des Multiplizierens abläuft, so kann diese Beschreibung multiplizieren. Man versuche das mal mit einem Buch! Allerdings wundert es niemanden mehr, daß Rechenanlagen multiplizieren können. Das Wundern beginnt, wenn Tätigkeiten beschrieben werden, die wesentlich zum Selbstverständnis des Menschen gehören, wie zum Beispiel das Lernen.

Die Lernfähigkeit betrifft den Menschen fundamental und wird daher (nicht nur) in Psychologie, Linguistik und Philosophie ausdifferenziert in verschiedenen Arbeitsgebieten untersucht. Die Kognitionspsychologie ebenso wie die Entwicklungspsychologie beschäftigen sich mit der Lernfähigkeit. Aus den vielen Inhalten, die gelernt werden, seien hier die Begriffe herausgegriffen: was sind Begriffe, wie werden sie gebildet? Von der einfachen Vorstellung, Begriffe könnten durch Merkmale erfaßt werden über die Prototypentheorie [17, 18] bis hin zu Begriffen als Theorien [13] wurden Beschreibungen für Begriffe und ihre Instanzen entwickelt. Ein Begriff (wie z.B. *Tasse*) deckt Instanzen ab (Einzeldinge, im Beispiel: einzelne Tassen), die alle ganz unterschiedlich sind bis auf die den Begriff definierenden Eigenschaften. Extensional ist ein Begriff durch seine Instanzen gegeben. Intensional ist ein Begriff durch seine Definition gegeben. Während die Kognitionspsychologie meist den Erwerb von (künstlichen) Begriffen unter Laborbedingungen untersucht und dafür Instanzen als Beispiele für den Begriff vorgibt, untersucht die Entwicklungspsychologie darüber hinaus, wie überhaupt erst einmal Mengen von Instanzen zusammengefaßt werden (Kategorisierung), bevor für diese Menge eine Definition gefunden wird. Man braucht nicht erst Goethes "Faust" zu zitieren, um vom Begriff auf das Wort zu kommen. Dem Begriff als Wortsemantik und seinen komplexen Strukturen (z.B. Ober- und Unterbegriffe) als implizite Verbindung von Sätzen zu Texten oder zu Dialogen wird in der Linguistik genau wie in der Psychologie eine herausragende Rolle zuerkannt. Wenn auch Chomsky bei seiner Lerntheorie sich insbesondere auf den

Grammatikerwerb bezieht [2], so behandeln Spracherwerbstheorien doch auch das Lernen von Wörtern mit ihrem Bezug auf Begriffe [1]. Schließlich hat die Philosophie, die ja die Lernfähigkeit in Hinblick auf Erkenntnis und Wahrheit aus dem individuellen mentalen Geschehen herausgehoben hat, eine wichtige Frage aufgeworfen, die die Begriffe und ihre Erlernbarkeit betrifft: worin sind sie verankert? Was muß gegeben sein, damit Begriffe gelernt werden können? Wie auch immer sie mental repräsentiert sind – ihre Repräsentation muß ihrerseits gelernt worden oder eben vorgegeben sein. Wie entkommt man aus dem infiniten Zirkel von Definitionen, die Begriffe verwenden, die ihrerseits definiert sind, wobei wiederum Begriffe verwendet werden ... ? Dies Problem des *symbol grounding* hat viele Philosophen beschäftigt [4, 5] und verzweigt sich hin zu so grundsätzlichen Fragen wie der Grenze zwischen (kognitivem) Lernen, biologischem Reifen und Angeborenem oder dem Spezialisierungsgrad der Lernfähigkeit für unterschiedliche Lerninhalte (ein universelles Lernverfahren oder für unterschiedliche Lernaufgaben spezialisierte Module?). Diese angedeuteten Verweise auf Aspekte der Lernfähigkeit, die – bezogen auf das Lernen von Begriffen – in Psychologie, Linguistik und Philosophie behandelt werden, dienen hier nur dazu, die zentrale Stellung des Phänomens in der *Menschenwelt* zu belegen.

Wenn nun ein Phänomen wie das Lernen operational beschrieben wird, so bringen wir es dadurch in die Maschinenwelt. Dort aber gibt es nichts nichts als das Beschriebene. Keine mitgedachten Kontexte, keine selbstverständlichen und deshalb nicht formalisierten Annahmen. In der Maschinenwelt ist nicht unbedingt das, was man meinte, sondern nur das, was man ausgedrückt hat. Unklarheiten und Unvollständigkeiten des eigenen Wissens über den Prozeß, den man beschreiben möchte, werden ebenso deutlich wie die partielle Unfähigkeit, sogar eigentlich Klares formal auszudrücken. Die Maschinenwelt zeigt uns unseren Wissensstand mit seinen Lücken und Fehlern. Die Maschinenwelt ist obendrein begrenzt durch ihre Ausdrucksmöglichkeiten und unsere Fähigkeit, mit ihnen geschickt umzugehen. Da das Wissen in der Maschinenwelt quasi eigenständig operiert, können wir es aus der Distanz betrachten – oder auch praktisch einsetzen.

1 Operationale Beschreibungen

Ich habe auf einer Rechenanlage ablaufende Programme als operationale Beschreibungen eingeführt. In diesem Abschnitt soll das etwas vertieft werden. Programme sind Beschreibungen in einer formalen Sprache. Jeder Ausdruck dieser Sprache hat eine Semantik, die durch Ausdrücke einer darunterliegenden Sprachschicht definiert ist. Letztlich werden alle Ausdrücke in Anweisungen der maschinennahen Sprache übersetzt. Diese elementaren Anweisungen werden von der Maschine ausgeführt. Programme sind also operational, weil sie von der Maschine interpretiert und ausgeführt werden. Da die Besonderheiten der Programmiersprache für die Beschreibung eines Prozesses keine Rolle spie-

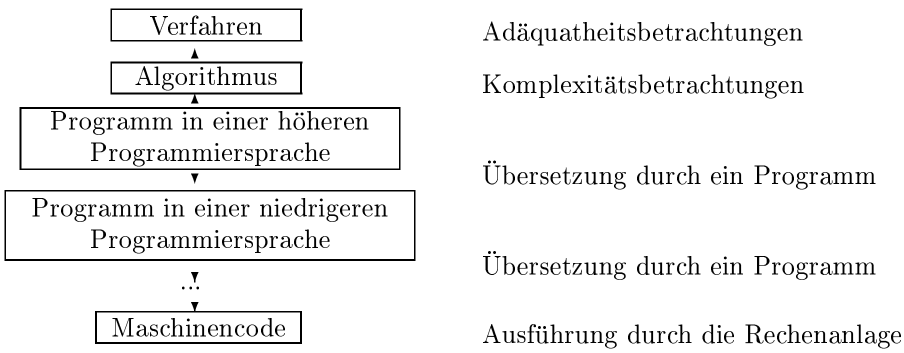

Abbildung 1: Abstraktionsebenen der Programmierung

len sollen, werden Programme ihrerseits beschrieben, z.B. in Form von Logik, in einer Spezifikationssprache oder einfach in natürlicher Sprache. Wenn wir von der syntaktischen Form und den konkreten Wörtern abstrahieren, so betrachten wir Algorithmen oder sogar Klassen von Algorithmen. Über diese abstrahierten Programme können wir Betrachtungen anstellen: ist es garantiert, daß der Algorithmus zu einem Ende kommt und anhält? Wie lange dauert die Ausfhrung des Algorithmus' im schlimmsten Falle? Wieviel Speicherplatz braucht der Algorithmus im schlimmsten Falle? Diese Fragen beziehen sich auf die Komplexität des Algorithmus' und sind unabhängig von der konkreten Programmierung. Wir wissen unabhängig von einem tatsächlichen Programm, ob ein Problem (die Aufgabenstellung für einen Algorithmus) sich überhaupt lösen läßt (berechenbar ist) und wie schwierig (komplex) die Aufgabe ist. Also reicht es für die allgemeine Analyse von Problemen aus, die abstrakte Ebene zu behandeln. Die abstrahierten Programme sind die eigentlichen operationalen Beschreibungen. Figur 1 zeigt schematisch die Abstraktionsebenen der Programmierung.

Das, was beschrieben wird, muß hinlänglich klar sein. Meist sind es einfache Vorgänge wie die Lohnbuchhaltung oder die Lagerhaltung eines Betriebs oder auch das Setzen von Texten für den Druck. Die betreffenden Lohnbuchhaltungs- oder Textverarbeitungsprogramme werden meist ohne eine Theorie über ihren Gegenstandsbereich erstellt [1]. In der Künstlichen Intelligenz werden jedoch meist solche Vorgänge beschrieben, über die in anderen Disziplinen Theorien existieren. Diese Theorien stellen das vorhandene Wissen dar, das nun operationalisiert werden soll. Dabei stellt sich oft heraus, daß die Theorien der anderen Disziplinen nicht so einfach operationalisierbar sind: sie sind unnötig speziell oder es fehlen gerade die ntigen Details, sie beziehen sich auf Zustände, wo die operationale Beschreibung es erlaubt, Prozesse zu formulieren, ... Deshalb

[1] Daß eine genaue Erfassung des Wissens von Setzern sehr sinnvoll für ein gutes Textverarbeitungsprogramm ist, hat Donald Knuth mit seinem Programm TeX gezeigt.

werden diese Theorien zwar sorgfältig beachtet, die Programme der Künstlichen Intelligenz stellen jedoch eine eigenständige Theorieentwicklung dar, die im besten Falle in den Diskurs der jeweiligen Disziplin eingebracht und dort auf ihre Adäquatheit hin untersucht wird. Ein wichtiger Vorteil der operationalen Beschreibungen besteht darin, daß festgestellt werden kann, ob sie tatsächlich das Verhalten beschreiben, das sie beschreiben sollen: man sieht nach, wie sich die Verfahren verhalten, indem man prft, ob sie wenigstens bezüglich des Ein-/Ausgabeverhaltens mit dem Beschriebenen übereinstimmen. Ein Verfahren für die Multiplikation sollte zumindest als Ausgabe genau das Produkt der eingegebenen Zahlen liefern. Wir mssen dann auf der Ebene des verwendeten Algorithmus prfen, ob unsere Testbeispiele Zufallstreffer waren oder ob der Algorithmus korrekt ist, also immer das richtige Ergebnis liefern wird. Wir wissen dann immer noch nicht, ob dies Verfahren das eleganteste, kognitiv adäquate Verfahren für die Multiplikation ist – immerhin ist es eine operationale Beschreibung. In derselben Weise können wir überprüfen, ob ein Lernverfahren tatsächlich eine Beschreibung (zumindest eines Aspekts) der Lernfähigkeit darstellt: lernt es aus einer Sequenz von Beispielen den Begriff, den auch Versuchspersonen in psychologischen Experimenten gelernt haben? Wenn ja, so bietet dies Lernverfahren uns ein Modell für das menschliche Lernen an. Ob es ein adäquates und elegantes Modell ist, entscheidet sich dann aufgrund metatheoretischer Kriterien. Immerhin bietet es im Gegensatz zu Modellen, die in Büchern beschrieben werden, neben dem Vorteil der Überprüfbarkeit noch zusätzlich den des praktischen Nutzens. Der Versuch, menschliche Lernprozesse zu beschreiben, hat uns lernende Programme eingebracht. Programme, die selbst lernen können, in denen also nicht die Lösung einer Aufgabe vorprogrammiert wurde, sondern stattdessen ein Verfahren, wie das Programm lernen kann, verschiedene Aufgaben zu lösen, ersparen viel Programmieraufwand. Selbst wenn sie also nicht elegant und adäquat sind, so sind sie zumindest in der Maschinenwelt sehr praktisch.

2 Lernaufgaben und Lernverfahren

Damit das Forschungsparadigma der Künstlichen Intelligenz nicht abstrakt bleibt, sollen in diesem Abschnitt Typen von Lernaufgaben beschrieben werden, die von Programmen gelöst werden. Es geht also um lernende Maschinen und ihre Verfahren. Dabei soll es hier lediglich um das Lernen von Begriffen gehen. Ich ordne die Lernaufgaben nach ihrer Schwierigkeit an, die wir komplexitätstheoretisch genau bestimmen können.

Die einfachste – und von der Psychologie bereits in frühen Arbeiten untersuchte – Lernaufgabe ist das Lernen eines Begriffs aus Beispielen, wobei die Begriffe einfach durch Merkmale definiert werden.

Lernen aus Beispielen

Gegeben: Beispiele, die durch Merkmale beschrieben und als Instanzen oder

Nicht-Instanzen eines Begriffs klassifiziert sind.

Lerne: Eine (möglichst kurze) Definition des Begriffs, so daß die Definition alle Instanzen abdeckt, aber keine Nicht-Instanzen.

Programme, die beschreiben, wie diese Lernaufgabe gelöst wird, sind als Produkte auf dem Markt erhltlich. Sie basieren auf statistischen Methoden. Die Verfahren bestehen aus zwei Algorithmen. Einer zählt nacheinander in einer sinnvollen Reihenfolge Hypothesen für Begriffsdefinitionen auf; der andere prüft bei jeder Hypothese, ob diese Definition alle Instanzen abdeckt und alle Nicht-Instanzen ausschließt. Diese Lernverfahren sind besonders dann nützlich, wenn entweder keine Begriffsdefinition bekannt ist oder wenn man die kürzeste Definition finden will. So können Experten oft nicht angeben, wie sie ihre Entscheidungen treffen, aber sehr wohl bei vorgelegten Beispielen die richtige Klassifikation angeben. Beispielsweise können Menschen handgeschriebene Buchstaben richtig erkennen, auch wenn die Schriftformen für einen Buchstaben bei verschiedenen Handschriften (und oft sogar bei derselben) stark voneinander abweichen. Ein Lernverfahren kann nun die Schriftform und den ihr von einem Menschen zugeordneten Buchstaben als Beispiele bekommen und daraufhin eine Regel ausgeben, die auch neuen, bisher nicht gesehenen Schriftformen je einen Buchstaben zuordnet. Das Programm hat dann gelernt, Buchstaben zu erkennen[2]. Diese Regel kann dann von einem Experten auf Plausibilität hin überprüft werden. Sie ist in den Ausdrücken formuliert, die als Merkmale ausgewählt wurden, und ihre Verständlichkeit hngt davon ab, wie geschickt die Merkmale gewählt wurden.

Schwieriger wird die Lernaufgabe, wenn zunächst die Beispiele in Mengen aufgeteilt werden müssen, wenn also noch keine Klassifikation vorgegeben ist.

Lernen aus Beobachtungen

Gegeben: Beobachtungen, die durch Merkmale beschrieben sind.

Lerne: Definitionen für Begriffe, die möglichst viele einander ähnliche Beobachtungen zusammenfassen, wobei die verschiedenen gewonnenen Begriffe untereinander möglichst unähnlich sein sollen.

Natürlich fällt gleich der enge Bezug zur Statistik (*clustering* Methoden) auf. Allerdings wird hier gefordert, daß die Beobachtungen nicht nur in verschiedene Mengen (*cluster*) aufgeteilt werden, sondern daß zu jeder Menge eine Begriffsdefinition angegeben wird, die wieder vom Experten des Sachbereichs verstanden und überprüft werden soll. Neue Beobachtungen werden dann gemäß der gefundenen Definitionen klassifiziert, können aber auch zu Veränderungen der Begriffe bzw. ihrer Definitionen führen.

[2] Auch neuronale Netze lösen diese einfache Lernaufgabe, wobei sie allerdings keine Regel ausgeben, die der Experte prüfen kann. Sie geben einfach für jede neue Schriftform einen Buchstaben aus.

Eine starke Vereinfachung der eben vorgestellten Lernaufgaben besteht bei den meisten Verfahren darin, daß sie elementare Merkmale verwenden, die ihrerseits nicht definiert werden. Damit wird ein prinzipieller Unterschied zwischen Merkmalen und Begriffen gemacht, was aber von der Sache her nicht angemessen ist. Denn jedes Merkmal kann auch als Begriff aufgefaßt werden. Andererseits kann auch jeder Begriff wiederum als Merkmal benutzt werden. So ist beispielsweise *menschlich* ein Merkmal und *Mensch* ein Begriff. Die tautologische Beziehung zwischen *Mensch* und *menschlich* kann aber von den einfacheren Lernverfahren nicht dargestellt werden. Sie können diesen Zusammenhang nicht beschreiben. Der Grund ist der Repräsentationsformalismus, der der Aussagenlogik entspricht. Erst wenn wir Formalismen wählen, die einer eingeschränkten Prädikatenlogik entsprechen, können wir in der Definition eines Begriffs als Merkmale wiederum Begriffe verwenden. Es können dann auch komplexere Begriffe gelernt werden. Begriffe wie *Großmutter* können gar nicht durch elementare Merkmale definiert werden. Sie sind ihrem Wesen nach relational. Sie werden durch eine Verkettung von *Mutter von-* und *Vater von-* Relationen definiert. Wählt man nun die Prädikatenlogik als Repräsentationsformalismus, so hält leider der Algorithmus, der feststellt, ob ein Beispiel von einer Definition abgedeckt wird, nicht unbedingt an. Wenn man Pech hat, rechnet er für alle Zeiten. Es werden deshalb Formalismen untersucht, die zwar noch die Darstellung von Relationen erlauben, aber dennoch so vereinfacht sind, daß die Überprüfung der Beispielabdeckung terminiert. Auch in der Psychologie wurde der einfache Merkmalsansatz kritisiert und es wurden komplexere Begriffe gefordert (Begriffe als Theorien). Prädikatenlogische Lernverfahren können diese Forderungen konkretisieren [6, 15, 19]. Sie sind dabei insofern detaillierter, als sie Annahmen explizieren, die das Lernen auch schwieriger Begriffe noch möglich machen. Psychologische Experimente, die diese Annahmen überprüfen, stehen noch aus.

Eine besonders schwierige Lernaufgabe ist die, alle für eine gegebene Menge von Beobachtungen gültigen Regeln zu finden. Dies entspricht dem Lernen nicht nur eines Begriffs, sondern der Entdeckung aller implizit vorhandenen Begriffe und ihrer Bezüge zueinander. Wenn zusätzlich gefordert ist, daß die Menge der Regeln minimal sein soll, handelt es sich um das Problem der Axiomatisierung: finde die Axiome, die genau die gegebenen wahren Sachverhalte beschreiben.

Wissensentdeckung

Gegeben: eine Menge von Beobachtungen.

Lerne: alle Regeln, die für die gegebenen Beobachtungen gültig sind. Dabei sollen die Regeln kürzer sein als die Menge der Beobachtungen. Es ist also nicht zulässig, die eingegebenen Aussagen einfach wieder auszugeben.

Verfahren der Wissensentdeckung ermöglichen es, Wissen, das implizit in Datenbanken verborgen ist, explizit zu machen. Wegen dieses praktischen Bedarfs sind Verfahren zur Wissensentdeckung ein Bereich, an dem zur Zeit international fieberhaft gearbeitet wird [14, 10, 3]. Sogar für die eingeschränkte Prädikatenlogik als Formalismus zur Repräsentation der Regeln gibt es bereits erste Verfahren [7, 9, 11].

Lernverfahren für alle diese Lernaufgaben (Lernen aus Beispielen, Lernen aus Beobachtungen, Wissensentdckung) bestehen aus Algorithmen für das Generieren von Hypothesen (möglichen Begriffsdefinitionen bzw. möglicherweise gültige Regeln) und solchen für das Testen von Hypothesen (Überprüfen, welche Beobachtungen von einer Hypothese abgedeckt werden). Sie geben ein Lernergebnis aus, das – wie gefordert – möglichst verständlich ist und folglich von Sachbereichsexperten auf Plausibilität geprüft werden kann. Das Lernergebnis kann dann von den Programmen genutzt werden, um neue, vorher nicht eingegebene Beobachtungen zu klassifizieren. Entscheidend für die Lernbarkeit ist der Formalismus, in dem das Lernergebnis dargestellt wird. Einfach lernbar sind Regeln in einem Formalismus, der aber nicht alles ausdrücken, so daß folglich das Lernverfahren auch nicht alles lernen kann. Während die frühen Lernverfahren der Künstlichen Intelligenz bereits zurückgeflossen sind in die psychologische Diskussion [12], steht dies für die neueren Verfahren noch aus. Allerdings gibt es bereits erste Anstrengungen zu interdisziplinärer Zusammenarbeit im Rahmen eines europäischen Projektes [16].

3 Lernende Roboter

Viele Ansätze zum maschinellen Lernen (lernende Maschinen) sind auf Eingaben durch Benutzer angewiesen, da der menschliche Benutzer den einzigen Kontakt der Maschine zur Wirklichkeit darstellt. Der Benutzer repräsentiert (d.h. formuliert) die Aspekte der Wirklichkeit, die ihn interessieren und gibt seine Beschreibung der Welt in das System ein. Es ist bereits schwieriger zu lernen, wenn die Beschreibung für einen anderen Zweck als das Lernverfahren erstellt wurde. Auch insofern ist die Wissensentdeckung in Datenbanken ein schwieriges Lernproblem: die Datenbank wurde für andere Zwecke als fr das Lernen angelegt und ihre Repräsentation ist vielleicht für Lernverfahren nicht gut geeignet. Aber auch in diesem Fall beruht der Zugang zur Wirklichkeit auf einer menschlichen Beschreibung. Erst Roboter nehmen unabhängig von einem Menschen mithilfe ihrer Sensoren die Wirklichkeit wahr. Daher stellt das Lernen von Robotern eine besondere Herausforderung dar. Die meisten lernenden Roboter optimieren entweder die Kontrolle von Handlungen wie z.B. das Greifen von Gegenständen (untere Ebene des Verhaltens) oder ihre Handlungspläne (obere Ebene). Sie lernen auf *einer* Ebene. Hier soll nun kurz ein Ansatz vorgestellt werden, bei dem die Verbindung von der unteren, direkt auf Wahrnehmung und Handlung bezogenen Ebene mit der oberen Ebene der Planung und Begriffe untersucht wird.

Eingangs wurde erwhnt, daß der Merkmalsansatz das Problem des *symbol grounding* in sich birgt. Wir wollen dies nun aufgreifen und sehen, wie es durch ein maschinelles Lernverfahren angegangen werden kann. Fodor schlug vor, die

Merkmale, mit denen wir Merkmale definieren, letztlich auf die Welt zurückzuführen [4]. Er spricht von "dem Prozeß, die Augen zu öffnen und zu sehen." Wrobel hat dies in seiner Argumentation gegen modische konnektionistische Ansätze, die angeblich nicht symbolisch sondern sub-symbolisch wären und damit das Problem des *symbol grounding* umgehen knnten, aufgegriffen und herausgestellt, daß man durch Wahrnehmung nicht den Symbolen entkommt[21]. Allerdings können die durch die Wahrnehmung empfangenen Symbole als elementar und angeboren betrachtet werden. Der infinite Regreß hört somit bei der Wahrnehmung auf. Die Verankerung der Symbole ist möglich. Die Verbindung von Basismerkmalen der Wahrnehmung zu abstrakteren Konzepten können wir in der Maschinenwelt besonders gut an Robotern untersuchen. Dabei brauchen wir uns nicht auf einfache Merkmale mit den bereits angeführten Kritikpunkten zu beschränken. Wir können relationale Begriffe als komplexe Merkmale zur Definition anderer Begriffe nutzen. Ja, wir können noch weiter gehen und einen Roboter aus seinen Handlungen und Wahrnehmungen Begriffe lernen lassen.

Begriffe sind letztlich in Handlung und Wahrnehmung begründet. Der Begriff der Tasse, der in psychologischen Studien untersucht wurde, definiert sich nicht allein durch den flachen Boden, die Öffnung oben, den Henkel, das Verhältnis von Höhe zu Breite des Behälters. Wesentlich ist, daß man aus einer Tasse trinken kann. Ist dies aus irgendeinem Grund nicht möglich, handelt es sich auch nicht um eine Tasse. Es wird niemals gelingen, alle Möglichkeiten, warum man aus einem Objekt nicht trinken kann, vollständig in der Definition dieses Objekts auszuschließen. Hat man gerade daran gedacht, daß der Henkel seitlich angebracht sein muß und nicht etwa über der Öffnung sein darf, so kann als nächstes ein Sieb als Boden auftreten; auch dann ist das Objekt keine Tasse. Der enscheidende Test, ob es sich um ein Trinkgefäß handelt (und alle Tassen sind Trinkgefäße), wird durch eine Handlung ausgeführt, deren Ergebnis sinnlich wahrgenommen wird. Hatten wir bisher in der Maschinenwelt ausschließlich Definitionen, die einen Begriff auf andere Begriffe (Merkmale) zurückführten, so ermöglicht uns ein Roboter, die Definition auch auf Handlungen und Wahrnehmungen zurückzuführen. Der Roboter erlaubt uns also, in der Maschinenwelt ein Fenster zur Wirklichkeit zu öffnen. Die Sensoren des Roboters messen etwas, das nicht programmiert wurde, sondern einfach existiert. Natürlich ist die Sensorik eines Roboters anders beschaffen als die eines Menschen und die Verarbeitung der Wahrnehmungen muß programmiert werden oder zumindest muß ein Lernverfahren programmiert werden, mit dem die Sensordaten verarbeitet werden.

In dem Projekt BLearn II, das von der europäischen Kommission gefördert wurde, ist es gelungen, handlungsbezogene Begriffe aus Sensordaten, Handlungsdaten und Benutzerklassifikationen zu lernen [8]. Ein an der Universität Karlsruhe entwickelter mobiler Roboter, der mit einfachen Ultraschallsensoren ausgetattet ist, fuhr durch einen Flur und an einer Türffnung vorbei oder durch eine Tür hindurch. Dabei zeichnete er die Abstandsmessungen der rund um den Roboter montierten Sensoren auf sowie die eigenen Bewegungen mit ihrer Richtung und Geschwindigkeit. Für jede solche Fahrt wurde nun von einem Menschen klassifiziert, ob es eine Fahrt an einer Tür vorbei, durch eine Tür hindurch oder nur entlang einer Wand war. Bei den Fahrten waren Flur und Raum unterschiedlich ausgestattet: mal stand ein Tisch neben der Tür, mal war ein Schrank an der Wand, mal nicht. Die Lernaufgabe war also ein Lernen aus Beispielen:

Gegeben: Fahrten mit Sensor- und Bewegungsdaten und eine Klassifikation der Fahrt nach den Begriffen *entlang einer Tür fahren, durch eine Tür fahren, entlang einer Wand fahren.*

Lerne: Definition der Begriffe in Ausdrücken von Sensor- und Bewegungsmustern.

Diese globale Lernaufgabe wurde in einzelne Schritte aufgeteilt. Der erste Schritt ist die Umwandlung einer Folge von Abstandsmessungen eines Roboters in ein Zeitintervall, in dem *mehr oder weniger* dasselbe Wahrnehmungsmuster gemessen wird. Das Ergebnis dieses Lernenprozesses kann dann für weitere Fahrten des Roboters genutzt werden. Während der Fahrt werden die Abstandsmessungen zu Folgen von Abstandslinien verarbeitet. Zum Beispiel wird festgestellt, daß von Zeitpunkt 0 bis Zeitpunkt 25 der Abstand *im wesentlichen gleich* blieb, von Zeitpunkt 26 bis Zeitpunkt 29 sich erheblich verringerte, sich von Zeitpunkt 30 bis 32 vergrößerte und dann wieder von Zeitpunkt 33 bis 41 konstant blieb. Solche Folgen von Abstnden werden für jeden der am Roboter angebrachten Sensoren festgestellt.

Die zweite Lernaufgabe ist dann, die Sequenzen der verschiedenen Sensoren zusammenzufassen und herauszufinden, welche Muster von welchen Sensoren zur Definition des Begriffes brauchbar sind. Dabei ist auch die zeitliche Beziehung wichtig. Beispielsweise sind es bei einer Fahrt durch die Tür natürlich gerade die seitlichen Sensoren, die fast gleichzeitig ein ähnliches Muster bei allen Fahrten wahrnehmen. Die tatsächlichen Zeitpunkte sind unwichtig. Sie hängen von der Größe der Tür und der Geschwindigkeit des Roboters ab. Wesentlich sind nur die zeitlichen Beziehungen *direkt vorher, vorher.* Bei der Fahrt entlang einer Tür sind nur die Messungen der Sensoren, die der Tür zugewandt sind, entscheidend. Das Lernverfahren findet die wichtigen Sensoren und die relevanten Muster. Es gibt Regeln aus, die aus den Wahrnehmungen und Bewegungen bei neuen Fahrten ableiten, ob der Roboter an einer Tür oder einer Wand entlang oder durch eine Tür fährt. Da nur die zeitlichen Relationen in den Definitionen vorkommen und nicht die tatsächlichen Zeitpunkte, können Türen unterschiedlicher Öffnungsbreite und mit unterschiedlichen Türrahmen als Türen erkannt werden. Etwas wird dann als *durch die Tür fahren* erkannt, wenn der Roboter tatsächlich hindurchfahren kann, so daß die hinteren Sensoren dann die Türöffnung wahrnehmen. Insofern ist eine für den Roboter zu kleine Tür für ihn keine Tür und ein Durchgang, der von Schränken gebildet wird, ist für ihn eine Tür. Hier zeigt sich die Abhängigkeit von der Körperlichkeit bei der Begriffsbildung. Die gelernten Regeln werden verwendet, um Handlungen zu planen und auszuführen. Hier werden also (gelernte) Beschreibungen nicht nur ausgeführt, indem neue Beschreibungen produziert werden, vielmehr bewegt sich die Maschine im Raum. Wenn dies auch kein prinzipieller Unterschied ist, so wird hier doch das Operationale der Beschreibungen besonders augenfällig.

Natürlich besteht der Hauptteil der Arbeit an lernenden Robotern bei dem eben beschriebenen Ansatz darin, schnelle Algorithmen für die Lernaufgaben und geeignete Repräsentationen für die Begriffe auf den verschiedenen Abstraktionsebenen zu finden. Es müssen Probleme der Parallelverarbeitung, der logischen Programmierung, der Datenhaltung und -übertragung gelöst werden. Diese Aspekte habe ich hier nicht weiter angesprochen, um eine der Zielsetzungen deutlich herauszustellen, die hinter diesen Arbeiten steht: die Verankerung von Begriffen in Handlung und Wahrnehmung mithilfe des Lernens.

4 Lernen und Lernenlassen

Die Maschinenwelt ist aus Sicht der Informatik eine Welt der Beschreibungen von Verhalten, wobei die Beschreibungen von den Maschinen selbst interpretiert werden. Die Interpretation ist immer Ausführung einer Handlung, sei es eine kognitive Handlung – z.B. zu einer Frage eine Antwort liefern –, sei es eine physische Handlung – z.B. durch eine Tür fahren. Es gibt nun (mindestens) zwei Gründe, warum wir Verhalten operational beschreiben. Der erste Grund ist der Erkenntnisgewinn, den wir uns von der Analyse der Beschreibungen erhoffen. Die Verhaltensbeschreibungen können als Beschreibungen auch menschlichen Verhaltens aufgefaßt werden. Dann kann experimentell überprüft werden, ob die Beschreibung zutrifft, ob also die Maschine ein menschenähnliches Verhalten zeigt. In dem Falle stimmen Beschreibung und Beschriebenes wenigstens in dem überein, was beobachtet werden kann. Die Beschreibung liefert also eine mögliche Hypothese über den nicht beobachtbaren inneren Vorgang. Die Beschreibung kann eine anregende Analogie für die Untersuchung menschlichen Verhaltens bieten. Sie kann Fragen aufwerfen und die Aufmerksamkeit auf Zusammenhänge lenken, die bei der Formalisierung deutlich wurden. Dieser Anspruch ist aber keineswegs zwingend. So muß beispielsweise die Beschreibung der Lernfähigkeit, die ein Lernverfahren bietet, keinesfalls als Modell menschlichen Lernens aufgefaßt werden. Sie bietet eine Formalisierung dessen, was wir als Lernen – bei wem oder was immer – auffassen wollen, letztlich also eine Definition des Begriffs *Lernen*.

Der zweite Grund ist die Nützlichkeit der Maschinen. Die operationalen Beschreibungen lassen Maschinen Verhalten zeigen, das von ihnen gewünscht wird. Sie erledigen Arbeiten für Menschen oder erleichtern Menschen ihre Arbeit. Dabei geht es nicht nur um die Erledigung von Aufgaben, die bereits von Menschen bearbeitet werden, sondern vor allem um zusätzliche Aufgaben, die man Menschen gar nicht zumuten würde oder die von Menschen nicht ausgeführt werden können. Die Lernfähigkeit macht Maschinen flexibler, schneller einsetzbar und kann Benutzer von Routinearbeiten entlasten. Während der erste Grund darauf verweist, daß wir Menschen (aus Beschreibungen von z.B. Lernvorgängen) lernen wollen, geht es bei dem zweiten Grund darum, daß wir uns Arbeit abnehmen lassen wollen. Dies kann auch heißen: wir lassen Maschinen lernen.

Literatur

[1] Susan Carey. The child as word learner. In M. Halle, J. Bresnan, and G. A. Miller, editors, *Linguistic theory and psychological reality.* MIT Press, Cambridge, MA, 1978.

[2] Noam Chomsky. *Reflexionen über die Sprache.* Suhrkamp, Frankfurt a.M., 1977.

[3] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining.* AAAI Press Series in Computer Science. A Bradford Book, The MIT Press, Cambridge Massachusetts, London England, 1996.

[4] J. A. Fodor. *The language of thought.* Thomas Y. Cromwell, New York, 1975.

[5] Stevan Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

[6] Jörg-Uwe Kietz and Katharina Morik. A polynomial approach to the constructive induction of structural knowledge. *Machine Learning Journal*, 14(2):193–217, 1994.

[7] Jörg-Uwe Kietz and Stefan Wrobel. Controlling the complexity of learning in logic through syntactic and task–oriented models. In Stephen Muggleton, editor, *Inductive Logic Programming*, chapter 16, pages 335–360. Academic Press, London, 1992.

[8] V. Klingspor, K. Morik, and A. Rieger. Learning concepts from sensor data of a mobile robot. *Machine Learning*, 23(2/3):305–332, 1996.

[9] Guido Lindner and Katharina Morik. Coupling a relational learning lagorithm with a database system. In Kodratoff, Nakhaeizadek, and Taylor, editors, *Statistics, Machine Learning, and Knowledge Discovery in Databases*, MLnet Familiarization Workshops, pages 163 – 168. MLnet, April 1995.

[10] Heikki Mannila. Aspects of data mining. In Yves Kodratoff, Gholemreza Nakhaeizadeh, and Charles Taylor, editors, *MLnet Workshop on Statistics,*

Machine Learning and Knowledge Discovery in Databases (ECML-95). Heraklion, Crete, Greece, 1995.

[11] Katharina Morik and Peter Brockhausen. A multistrategy approach to relational knowledge discovery in databases. In Ryszard S. Michalski and Janusz Wnek, editors, *Proceedings of the Third International Workshop on Multistrategy Learning (MSL-96)*, Palo Alto, May 1996. AAAI Press.

[12] Gregory L. Murphy. Theories and concept formation. In Iven Van Mechelen, James Hampton, Ryszard S. Michalski, and Peter Theuns, editors, *Categories and Concepts. Theoretical Views and Inductive Data Analysis*, Cognitive Science Series, chapter 7, pages 173–202. Academic Press, London, 1993.

[13] Gregory L. Murphy and Douglas L. Medin. The role of theories in conceptual coherence. *Psychological Review*, 92(3):289–316, July 1985.

[14] Gregory Piatetsky-Shapiro, William Frawley, and Christopher Matheus. Knowledge discovery in databases - an overview. In Gregory Piatetsky-Shapiro and William Frawley, editors, *Knowledge Discovery in Databases*, pages 1–27, Menlo Park, 1991. AAAI, AAAI Press.

[15] J.R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5(3):239–266, 1990.

[16] Peter Reimann and Hans Spada, editors. *Learning in Humans and Machines - Towards an Interdisciplinary Learning Science*. Elsevier, London, 1996.

[17] E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104:192–233, 1975.

[18] E. Rosch. Principles of categorization. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Erlbaum, Hillsdale, NJ, 1978.

[19] Ehud Y. Shapiro. *Algorithmic Program Debugging*. ACM Distinguished Doctoral Dissertations. The MIT Press, Cambridge, Mass., 1983.

[20] Sherry Turkle. *The Second Self. Computers and the Human Spirit*. Deutsch als Die Wunschmaschine- vom Entstehen der Computerkultur bei Rowohlt. Simon and Schuster, New York, 1984.

[21] S. Wrobel. *Concept Formation and Knowledge Revision*. Kluwer Academic Publishers, 1994.